

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2002, 00:47:10 ; Search time 2103.99 Seconds
(without alignments)
15028.587 Million cell updates/sec

Title: US-09-786-136-4
Perfect score: 1511
Sequence: 1 tccagagtcgaattctg.....atcctatgcaataaagtgcga 1511

ng table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_hg:.*
2: gb_hg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vl:.*
15: em_ba:.*
16: em_fun:.*
17: em_in:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*
29: em_vl:.*
30: em_htg_hum:.*
31: em_htg_inv:.*
32: em_htg_inv:.*
33: em_htg_inv:.*

Cite hits 1-97?

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	100	100	1	1

1	1294.4	85.7	1903	9	AB051125
2	1268.8	84.0	3319	9	BC001200
3	1217.8	80.6	151032	2	AC061705
4	1217.8	80.6	229586	2	AC048331
5	1214.6	80.4	152623	2	AC024530
6	1211.4	80.2	298408	2	AC092931
7	1110.4	73.5	2814	9	AB047829
8	471.8	31.2	198696	2	AC087898
9	471.8	31.2	199359	2	AC074027
10	471.8	31.2	219564	2	AC073708
11	271.8	18.0	173803	9	CNS01DV2
12	271.8	18.0	208953	9	CNS01RH3
13	103.4	6.8	80767	9	HS745E8
14	103.4	6.8	146722	2	AL135254
15	100.4	6.6	2116	9	AK057346
16	95	6.3	61503	2	AC107845
17	84.8	5.6	125020	9	AF429315
18	70.8	4.7	125020	9	AF429315
19	65.6	4.3	168554	2	AC096863
20	63.8	4.2	207064	2	AC106680
21	62.6	4.1	197669	2	AC084064
22	61.2	4.1	176370	2	AC105829
23	60.8	4.0	10732	6	E32986
24	59	3.9	7218	6	I66494
25	57.6	3.8	205914	2	AC027682
26	57	3.8	55021	2	AC107821
27	57	3.8	55536	2	AC090310
28	54.8	3.6	129600	2	AC097133
29	54.8	3.6	219952	2	AC084804
30	54.4	3.6	176877	2	AP003902
31	54.4	3.6	298166	2	AC087563
32	54.4	3.6	300695	2	AC079431
33	54.2	3.6	9389	1	STE250581
34	54.2	3.6	166341	2	AC013381
35	54	3.6	4233	9	HSU75308
36	54	3.6	90935	2	AP003917
37	54	3.6	161307	2	AC084058
38	54	3.6	176986	2	AC108671
39	53.8	3.6	92310	2	AC108380
40	53.8	3.6	94914	2	AC105592
41	53.2	3.5	303091	2	AC084799
42	52.8	3.5	104341	2	AC099456
43	52.6	3.5	82815	2	AC062001
44	52.2	3.5	85434	2	AC066610
45	52.2	3.5	221852	2	AC069214

ALIGNMENTS

RESULT 1

LOCUS AB051125

DEFINITION Macaca fascicularis brain cDNA, clone:QcCE-20281.

ACCESSION AB051125

VERSION AB051125.1 GI:11231112

KEYWORDS fis (full insert sequence).

SOURCE Macaca fascicularis adult male cerebellum cortex cDNA to mRNA, clone.lib:macaque brain cDNA library QcCE clone:QcCE-20281.

ORGANISM Macaca fascicularis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;

REFERENCE 1 (bases 1 to 1903)

AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirata,M., Suto,Y., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.

TITLE Assignment of 118 novel cDNAs of cynomolgus monkey brain to human chromosomes

JOURNAL Gene 275 (1), 31-37 (2001)

MEDLINE 21458551

REFERENCE 2 (bases 1 to 1903)

AUTHORS Hashimoto,K., Osada,N., Kusuda,J. and Sugano,S.

TITLE Direct Submission

AB051125 Macaca fa
BC001200 Homo sapi
AC061705 Homo sapi
AC048331 Homo sapi
AC024530 Homo sapi
AC092931 Homo sapi
AB047829 Macaca fa
AC087898 Mus muscu
AC074027 Mus muscu
AC073708 Mus muscu
AL133448 Human chr
AL160314 Human chr
AL1020398 Homo sapi
AL139254 Homo sapi
AK057346 Homo sapi
AC107845 Mus muscu
AF429315 Homo sapi
AF429315 Homo sapi
AC096863 Mus muscu
AC106680 Rattus no
AC084064 Homo sapi
AC105829 Rattus no
E32986 Gene encodi
I66494 Sequence 14
AC027682 Homo sapi
AC107821 Mus muscu
AC090310 Homo sapi
AC097133 Rattus no
AC084804 Mus muscu
AP003902 Oryza sat
AC087563 Homo sapi
AC079431 Mus muscu
AJ250581 Streptomy
AC013381 Homo sapi
U75308 Human TBP-a
AP003917 Oryza sat
AC084058 Homo sapi
AC108671 Homo sapi
AC108380 Pan trogl
AC105592 Rattus no
AC084799 Mus muscu
AC099456 Rattus no
AC062001 Homo sapi
AC066610 Homo sapi
AC069214 Homo sapi

JOURNAL

Submitted (14-NOV-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111 (ex.2120), Fax: 81-3-5285-1181)

COMMENT

Lab host: T0P10

Vector: pME18S-FL3 (Acc.No. AB009864)

R. Site1: DraIII (CAGCTGTGTG)

R. Site2: DraIII (CAGCTGTGTG)

Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing

(5' end primer [CTCTGTCTTAAAGCTGCG];

3' end primer [CGACCTGCAGCTCGACACA]);

Location/Qualifiers

1. .1903

/organism="Macaca fascicularis"

/db_xref="taxon:9541"

/clone="QcCE-20281"

/sex="male"

/tissue_type="cerebellum cortex"

/clone_lib="macaque brain cDNA library QcCE"

/dev_stage="adult"

809. .1591

/codon_start=1

/product="hypothetical protein"

/protein_id="BAB18151.1"

/db_xref="GI:11231113"

/translation="MIPTAGLOGSPAGAWDSNRNKSICALGDAATPMEGPRCPPPR
SPRLSLRRHRLCPDLGOANNSPDIDHVLPLVLOEAPCSPLDAPFCAAVRISO
ERLCRASPNVHRSLSPTASLPLWLLPGVGQDSATACSPSPSGSGPQVDS
RGSDTEASEGEGGLGGDLRGRTWATAVALWLEHRCAAAFGEWELFAARADCLWLA
QHLPGDLDAALKAARGLFLLLRHWADGNLQHLHCYSPANV"

374 a 597 c 537 g 395 t

BASE COUNT

ORIGIN

Query Match 85.7%; Score 1294.4; DB 9; Length 1903;

Best Local Similarity 91.1%; Pred. No. 8.9e-224;

Matches 1436; Conservative 0; Mismatches 221; Indels 69; Gaps 3;

QY 1 tccagagtgtgaattcttggcagccagggggtcttctgtctgtatgggtgaagatt 60

Db 299 TCAGAGTGTCTGAA--TCTGGGGCAGCCAGGGGGCCCTTGTCTGCTACCGGTTGAAGGTT 357

QY 61 cagcctgatacacttcttgcataagttgggagtgagctggttcttggccttagcaagc 120

Db 358 CAGCGTGACCACTTCTTGTCAAAAGTCGGGAGTCTCAGCTGGTTCCTGGCCCTAGCAAGGC 417

QY 121 aactactctgttaagtcgataagagggagagctgggagaaacactgcaatttttcta 180

Db 418 AATCTACTCTGTTAAAGTCGATAGAGGAGAAAGCTGGGAGAAATAATGCAATTTTTCTA 477

QY 181 tgaacctctgtagaggtggtggtggggtccacagactgctctgtgttccacagg 240

Db 478 TGAGCCCTCTGTAGAGGGTGGGTGAAGGGTTCACAGACCCTCTTGGTTCCACAGG 537

QY 241 ccattgcccggaggtgctggtctcgagcccttcagacaagtaagtcagctctgccccttc 300

Db 538 CCATGCCGGAGGTGCTGGCTTCGACCCCTTCACAGAGTAAGTTCAGCTCTGCCCTTC 597

QY 301 ctgcttaacttgcctgtagctgtgatctactactagggaggttctgctggggccct 360

Db 598 CTGCTTCACTTGCCTGTAGCTGTGATGTACCACTAGGAGGTCCTGCTGGGTCT 657

QY 361 gcagggtgcagctcagagccctagccccaggaacccctctgctctcagagcca 420

Db 1736 CCCCCACAAGAGTGCCTTGCCTCTCTCCGCTCTCTCCACCCACACACCCCTCCC 1795

Db 658 GCAGGTGTGCAGCTCAGAGCCAGCTGACCCCCCAGGAACCCCTCTCTGCCGCTCACAGCCG 717

QY 421 tctagatgcagctcctctgcccactgtgtgtctactctaa----- 459

Db 718 TCTAGATGCAGCTCCTCTGCTACTGTGTCTACTCTAAAGGTAGGGAAGGGTAGGGGC 777

QY 460 -----aggacttcagagagg 474

Db 778 ACTTGGACTTAGAGACCCCGCTGGCACTGATGATCCCCACTTCAGAGGAGG 837

QY 475 ctctcagcagcgctggtgactcggaataatggcaatcccaagcgtgtgttggggga 534

Db 838 CTCCTCACAGCGCTGGGACTGGACAGAAATGGCAACTCCCAAGTGTGCTTTGGGGGA 897

QY 535 ccttgcaactccacggaaggtcctcgccgcaactccccctcctcctcctcctcctcag 594

Db 898 CGCTGCCACTCCCATGGAAGTCTCTGCTGCCACCTCCCGCTTCTCCTCTCGGCTAG 957

QY 595 catggcgccgctcacaaactctgtagccttgaccgcccgggagcccaaacagtgagg 654

Db 958 CTTGGGCGCGCTCACAAACTCTGTAGACCTGAGCTGGGCGAGGCCAACACAGTGAAGG 1017

QY 655 cagcagaccatgactacctgccccttgggtgcgctgcagagagcaccaggtccttccgct 714

Db 1018 CATCGACCATGACTACCTTGCCTTGGTGGCTGCGAGGAGCACAGGCTCCTTCCGCT 1077

QY 715 ggaagcgccttctgcccgcgtgtgcctcgcagagcgctccttgcctggtgcctgccc 774

Db 1078 GGACGCGCCCTTCTGTGCGCTGTGCGCATCTCGCAGGAGCGCTCTGCGCGGCTGCGC 1137

QY 775 ctttgctgacagcgccagcctcagcccaactcagcccaactcagccctggtgggcaactct 834

Db 1138 CTTTCCCTGTCACCGTGCAGCCTCAGCCCACTCGGCTCATTTGCCCTGGGCACTTCT 1197

QY 835 gggccctggtgtggcaggtgacagtgccagcgctcctcctgagccgctcccccaactc 894

Db 1198 AGCCCTCGTGTGTGCTAGGTGACAGTGCACAGCTTCTGACGCGCTGCCCCAGCTC 1257

QY 895 gggctctgaggggcagcccagggtggacagtgggcggtcagacacagagggctccga 954

Db 1258 GGGCTCCGAGGGTCCGGGCCAGGTGGACAGTGGGGGGCTCAGATACCGAGGCTCGGA 1317

QY 955 tggggcggaagggctggcgccagcactcggggcgagccctggccactgcccagtc 1014

Db 1318 GGGGCGGAAGGGTGGGCGCACCGACCTCGGGGCGCGGACCTGGGCACTTGCCTGGC 1377

QY 1015 actcctcctggttgagcaccagtcgcccgtcctctgacagtgaggagctgacagcgc 1074

Db 1378 GCTGCCCTGGCTAGAGCACCGATGGCTGCTGCTTTGGCAGTGGGAACTGACAGCAGC 1437

QY 1075 caaggctgattgctggctggggccagcacttgctgacggccttgacctggcgccct 1134

Db 1438 CAAGGCTGATTGCTGGCTGGGCGCCAGCACTTGCTGACGGCTTGATCTGGCGGCT 1497

QY 1135 caaggccgagcccgagggctctctcctgctactgcccactgggacaaacacccctgagct 1194

Db 1498 CAAGGCCGACGCCGGGGGCTCTTCTGCTACTGCGGCACCTGGGACCAACACCTGCACT 1557

QY 1195 acacctgctgtgtacagccagcgaacgtgtaaggtcgccctcctgctgtggtgctg 1254

Db 1558 ACACCTGCTGTGTACAGCCCAACCAACGTGAAGGTGTCGCGCTGCTGTGGTGG 1617

QY 1255 gggcccccacaaacacactcaagtcactcggccagggctggcctcttgggtgggaaa 1314

Db 1618 CGCCCCACCAACACACT--AGACACTGCGCCCGCCAGGACTGCGCTCTTGGTGGTGGGAAG 1675

QY 1315 gtgtaggtgtgtgcagcctgtcccccactgttcttactcctcctcctcctcctcctt 1374

Db 1676 GTGTAGGTGTGTGCAGGCTGTCCGCCACTGCTTCTACTCTCCCTCCCTTAGAGGCTCTTG 1735

QY 1375 cccccacaaaagtgctgctgctgtctcctcctcctcctcctcctcctcctcctcctc 1434

Db 1736 CCCCCACAAGAGTGCCTTGCCTCTCTCCGCTCTCTCCACCCACACACCCCTCCC 1795

Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 70% of reads
 Assembly program: Phrap; version 0.990329 First call to
 findPhrapList
 Consensus quality: 230423 bases at least Q40
 Consensus quality: 245348 bases at least Q30
 Consensus quality: 254396 bases at least Q20
 Estimated insert size: 236955; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 23 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved

1 102971: contig of 102971 bp in length
 102972 103071: gap of unknown length
 103072 138452: contig of 35381 bp in length
 138453 138552: gap of unknown length
 138553 157218: contig of 18666 bp in length
 157219 157318: gap of unknown length
 157319 165594: contig of 8276 bp in length
 165595 165694: gap of unknown length
 165695 170674: contig of 4979 bp in length
 170675 170773: gap of unknown length
 170774 177554: contig of 6881 bp in length
 177555 181770: contig of 4016 bp in length
 181771 181870: gap of unknown length
 181871 185396: contig of 3426 bp in length
 185397 185396: gap of unknown length
 185397 190050: contig of 4554 bp in length
 190051 193028: contig of 2978 bp in length
 193029 193128: gap of unknown length
 193129 195613: contig of 2485 bp in length
 195614 199136: contig of 3423 bp in length
 199137 199236: gap of unknown length
 199237 202491: contig of 3255 bp in length
 202492 202591: gap of unknown length
 202592 205157: contig of 2566 bp in length
 205158 205257: gap of unknown length
 205258 207859: contig of 2612 bp in length
 207860 207959: gap of unknown length
 210484 210584: contig of 2515 bp in length
 210585 213846: contig of 3262 bp in length
 213847 213946: gap of unknown length
 213947 217298: contig of 3352 bp in length
 217299 217398: gap of unknown length
 217399 220482: contig of 3084 bp in length
 220483 220582: gap of unknown length
 220583 223048: contig of 2466 bp in length
 223049 225148: gap of unknown length
 225149 225384: gap of unknown length
 225385 227397: contig of 2012 bp in length
 227397 227497: gap of unknown length
 227497 229586: contig of 2090 bp in length.

Location/Qualifiers
 1. 229586
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-48102"

FEATURES
 source

AC048331 229586 bp DNA linear HTG 31-JAN-2002
 Homo sapiens chromosome 3 clone RP11-48102, WORKING DRAFT SEQUENCE.
 23 unordered pieces.
 AC048331
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 229586)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Albrooks, S.L., Amarante, R.C., Are, J.R., Ayale, M., Banks, T.,
 Barabara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Bathorne, S.K., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsif, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, N., Neal, Z., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,
 Weinstein, G., and Gibbs, R.
 Direct Submission

2 (bases 1 to 229586)
 Worley, K.C.
 Direct Submission
 Submitted (14-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jan 31, 2002 this sequence version replaced gi:17974902.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HANQ
 Center clone name: RP11-48102
 ----- Summary Statistics

	BASE COUNT	56229 a	57475 c	54442 g	59219 t	2221 others
ORIGIN						

Query Match 80.6%; Score 1217.8; DB 2; Length 229586;
Best Local Similarity 85.2%; Pred. NO. 3.7e-210;
Matches 1508; Conservative 0; Mismatches 2; Indels 259; Gaps 3;

Qy 1 tccagagtgctgaatttctgggcagccaggggcctctgctctgctatgggttgaagatt 60
|||||
Db 83695 TCCAGAGTGTGAAATTTCTGGGCAGCCAGGGGGCTCTTGCTCTGCTATGGGTGAAGATT 83754

Qy 61 cagcctgataactctcttgctcaaaagttagggagctctgagctggtctctgggcttagcaaggc 120
|||||
83755 CAGCGTGATCACTTCCTGTCAAAAGTTGGGGGACTCTGAGCTGCTCTGGGCCCTAGCAAGGC 83814
|||||

121 aatcactctgttaagtcgatagaggagagctggggagacactgcaatttttcta 180
|||||
83815 AATCTACTCTCTCTTAATAACTCGGATAGAGGCGAAGCTGGCGGCGAGCACTTCGCAATTTTCTCTA 83874

[illegible]

22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 5

DB 83935 CCATGCCCGAGGTGCTGGCTTCGAGGCCCTTCAGACAAAGTAAGGTCAGCTCTGCCCCCTC 83991

QY 301 ctgcttcacttgcctgtagctgtggtgctactactagggaggtcctgcctggggccct 360

Db	83995	CTGCTTCAC	TGCCCCG	TAGCTG	TGGATG	CTACTACT	AGGAGG	TCTG	CC	TGGGGCCCT	84054
Qy	361	gcaggtgtg	agagct	-----							374

Qy	375	-----cagagccgct 385
Db	84055	GCAGGTGCAGCTCAGGTAGGGGCCCTCTCTCTGGTGACCTCTCAGGTGCAGTCTCACTT 84114

Db 84115 CACTGGCCACTCTCTCTGGACAAACCTGACTCTCTGTCCCATGTGCCCTGCAGAGCCCGCT 84174
Qy 386 gagccccaggaaacccctctgcctctcacagccatctagatgcagctctctctgcccact 445

Db	84175	GAGCCCCCAGGAACCCCTCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCACT	84234
Qy	446	gttggtctactctaa-----	459

84235 GTTGTCTACTCTAAAGGTAACACCCAAAGGTAGGGGAAGGTAGGGGCACCTTGGGCTTAG 84294
|||||
460 -----aggacttcagaagacctctccacagg 486

Db 84295 AGACGCCCCCTGGCACTGATGATCCCACTGCAGGACTTCAGAGAGSCTCTCCAGCAGG 84354

Db 84355 GCCTGGGACTCGGACCAAAATGGCACTCCAAAGCGTGCTTTGGGGGACCTTGCCACTCC 84414

QY 347 CACGGAAGGTCCTCGCGGCCACCTCCCGTCTCCCTGTCGCTCAGCATGGCGCGCG 606
|||||
Db 84415 CACGGAAGGTCCTCGCGGCCACCTCCCGTCTCCCTGTCGCTCAGCATGGCGCGCG 84474

Qy 607 tcacaaactctgtagcctgaccggcgccagccaaacagtgaaagcagcagccactga 666
|||||
Db 84475 TCACAACTCTTAGCCTGACCGGGCCGCGCCCAACAACAGTGAGGCAGGCCACCATGA 84534

Qy 667 ctaccgacctt----- 678
 |||||
 Db 84535 CTACCTGCCCTTGGTGAGGACTCGGAGGTGGAGGTGGTGCCTCGGGCCGGCGCGTG 84594

Qy 679 -----ggtggcgtgcaggagccacg 701
|||
Db 84595 TTTCAGTCGCTTCTCCCCCACCCTCTCTCCTCAGGTGGGCTGCAGGAGGCACACG 84654

Qy 702 gctccttcgcctggacgcgcacctctctgcgcgcctgtgcgcattctgcagagagcgctct 761

Db 8455 GTCCTTCGGCCCTGACAGGCCCTTCTGCGCGCTGTGCGCATCTCGCAGGAGCGCCTCT 84714

Qy 762 gccgtgctgcgcctttgcggtgcacggcgcagcctcagcccaacctggcctcattgc 821
|||||
Db 84715 gccgtgctgcgcctttgcggtgcacggcgcagcctcagcccaacctggcctcattgc 84774
|||||

[illegible]

Qy 882 cgtccccagctcgggctctgaggggccaggccaggctggacagtggggggggtcagaca 941

DD 84 653 CGTCCCCCAGATCTGGGGCTCTGAGAGGGGCCAGGGCTGGGGGCGGGGCTCAGACA 84894
 Qy 942 ccagggctccgatggggcggaaggctggcggcacgacctgcggggccggacctggg 1001

DB 84895 CCAGGGCTCCGAGGGGGCGAAGGGTGGCGCCACCGGACCTGGG 8495A
QY 1002 ccaactgcgtagcactgcctlggtggagaccgacccgatgcgccgtgccttcgacagtggg 1061

Accession	Sequence	Length
Db 84955	CCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGCGCCCGTGCCTTCGACGAGTGGG	85014
Qy 1062	aactgacagcgggccaaggctgattgctggctgcgggggccagcacttgctgtgacggccttg	1121

Db 85015 AACTGACAGCGGCCAAGGCTGATTGTGGCTGCGGGCCAGCACTTGCCTGACGGCCTTG 85074

Qy 1122 acctggcgccctcaaggccgcagcccgaggctcttctctgctactgcgccactggggacc 1181

Db 85075 ACCTGGCGCCCTCAAGGCGCAGCCGAGGCTCTTCTGCTACTGCACCACATGGGACC 85134
Qy 1182 aaacactgcagctacacactgctgtgtctacagcccgagaaagtgtgaaggctgccccctg 1241

Db 85135 AAAAACTGCAGCTACACCTGCTGTGTACAGCCAGCAACGTTGAAGGCTGCCCCCTG 85194

Qy 1242 ctgcttggagctggcccccaccccaacacacactcaagtcactgcgcccgaaggctgagcctct 1301

		CTGCTTGGGTGGCGCCCAACACACTCAAGTCAC TGCCGCCAAGGCTGGCCTCT	85254
Db	85195		
Ov	1302	tggctactgagaaagtataggctggtaccagcgcattcccccaactgcttcttacttcctcc	1361

Db 85255 TGGTGTGGGAAGTGTAGGTGGTGCAGCTGTCCCCACGTCTTACTCCCTCCC 85314

Db 85315 TAGAGCCCTCTTGC CCCCAAAAAGTGCGCTGTGCTCTTCCTCTCCTCCCACCC 85374

QY 1422 CACCCACATCCCCCATCCCTCTGAGTCCCTGCAACACAGTGGAGGGTAGAGAGCCA 1481

Db 85375 CACTCACATCCCCCTCCATCCTCTGAGTCCCTGCAACACAGTGGAGGGTAGAGAGCCA 85434

Qy 1482 cagtccccaaatccctatgcaataaagtgc 1510
 |||||
 Db 85435 CAGTCCCCAAATCCTATGCAATAAAGTGC 85463

LOCUS	AC024530	159623 bp	DNA	17-SEP-84	TIME 14 MAR 85
RESULT 5					
AC024530/C					

DEFINITION Homo sapiens chromosome 4 clone RP11-131020 map 4, *** SEQUENCING IN PROGRESS ***
IN PROGRESS ***
ACCESSION AC024530

VERSION	AC024530.3	GI: 239669
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
Birren B., Linton J., Nushbaum C. and Lander F.
1 (bases 1 to 152623)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 152623)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

TITLE	Homo sapiens chromosome 4, clone RP11-131020
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 152623)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Glende, S., Coyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Iliev, T., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Laroque, K., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

LE

RNAL

COMMENT

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 14, 2000 this sequence version replaced gi:7229935.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7361

Center clone name: 131_O_20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

1 1156: contig of 1156 bp in length
1157 1256: gap of 100 bp
1257 2440: contig of 1184 bp in length
2441 2540: gap of 100 bp
2541 3807: contig of 1267 bp in length
3808 3907: gap of 100 bp
3908 5016: contig of 1109 bp in length
5017 5116: gap of 100 bp
5117 6229: contig of 1113 bp in length
6230 6329: gap of 100 bp
6330 7466: contig of 1137 bp in length
7467 7566: gap of 100 bp
7567 8835: contig of 1269 bp in length
8836 8935: gap of 100 bp
8936 10247: contig of 1312 bp in length
10248 10347: gap of 100 bp
10348 11428: contig of 1081 bp in length
11429 11528: gap of 100 bp
11529 12880: contig of 1352 bp in length
12881 12980: gap of 100 bp
12981 13991: contig of 1011 bp in length
13992 14091: gap of 100 bp
14092 15759: contig of 1668 bp in length
15760 15859: gap of 100 bp
15860 17156: contig of 1297 bp in length
17157 17256: gap of 100 bp
17257 18728: contig of 1472 bp in length
18729 18828: gap of 100 bp
18829 20169: contig of 1341 bp in length
20170 20269: gap of 100 bp
20270 21411: contig of 1142 bp in length
21412 21511: gap of 100 bp
21512 22964: contig of 1453 bp in length
22965 23064: gap of 100 bp
23065 25024: contig of 1960 bp in length
25025 25124: gap of 100 bp
25125 27388: contig of 2264 bp in length
27389 27488: gap of 100 bp
27489 28881: contig of 1393 bp in length
28882 28981: gap of 100 bp
28982 30874: contig of 1893 bp in length
30875 30974: gap of 100 bp
30975 33014: contig of 2040 bp in length
33015 33114: gap of 100 bp
33115 34935: contig of 1821 bp in length
34936 35035: gap of 100 bp
35036 37222: contig of 2187 bp in length
37223 37322: gap of 100 bp
37323 39483: contig of 2161 bp in length
39484 39583: gap of 100 bp
39584 41860: contig of 2277 bp in length
41861 41960: gap of 100 bp
41961 44448: contig of 2488 bp in length
44449 44548: gap of 100 bp
44549 46452: contig of 1904 bp in length
46453 46552: gap of 100 bp
46553 49889: contig of 3337 bp in length
49890 49989: gap of 100 bp
49990 52994: contig of 3005 bp in length
52995 53094: gap of 100 bp
53095 56588: contig of 3494 bp in length
56589 56688: gap of 100 bp
56689 59956: contig of 3268 bp in length
59957 60056: gap of 100 bp
60057 63904: contig of 3848 bp in length
63905 64004: gap of 100 bp
64005 67438: contig of 3434 bp in length
67439 67538: gap of 100 bp
67539 70684: contig of 3146 bp in length
70685 70784: gap of 100 bp
70785 75479: contig of 4695 bp in length
75480 75579: gap of 100 bp
75580 79243: contig of 3664 bp in length
79244 79343: gap of 100 bp
79344 82983: contig of 3640 bp in length
82984 83083: gap of 100 bp
83084 87910: contig of 4827 bp in length
87911 88010: gap of 100 bp
88011 94315: contig of 6305 bp in length
94316 94415: gap of 100 bp
94416 99313: contig of 4898 bp in length
99314 99413: gap of 100 bp
99414 104977: contig of 5564 bp in length
104978 105077: gap of 100 bp
105078 112571: contig of 7494 bp in length
112572 112671: gap of 100 bp
112672 118896: contig of 6225 bp in length
118897 118996: gap of 100 bp
118997 125479: contig of 6483 bp in length
125480 125579: gap of 100 bp
125580 130800: contig of 5221 bp in length
130801 130900: gap of 100 bp
130901 137670: contig of 6770 bp in length
137671 137770: gap of 100 bp
137771 145779: contig of 8009 bp in length
145780 145879: gap of 100 bp
145880 152623: contig of 6744 bp in length.
Location/Qualifiers
1. .152623
/db_xref="Homo sapiens"
/db_xref="taxon:9606"

```

FEATURES

source

[REDACTED]

COMMENT

*	1	33307:	contig of 33307 bp in length
*	33308	33407:	gap of unknown length
*	33408	55576:	contig of 22159 bp in length
*	55577	55676:	gap of unknown length
*	71624	71524:	contig of 15848 bp in length
*	71625	71624:	gap of unknown length
*	84811:	84811:	contig of 13187 bp in length
*	84812	84911:	gap of unknown length
*	84912	96841:	contig of 11930 bp in length
*	96842	96941:	gap of unknown length
*	96942	107204:	contig of 10263 bp in length
*	107205	107304:	gap of unknown length
*	107305	119046:	contig of 11742 bp in length
*	119047	119146:	gap of unknown length
*	119147	128812:	contig of 9666 bp in length
*	128813	128912:	gap of unknown length
*	128913	129236:	contig of 10324 bp in length
*	129237	139335:	gap of unknown length
*	139337	148308:	contig of 8972 bp in length
*	148309	148408:	gap of unknown length
*	148409	156373:	contig of 7965 bp in length
*	156374	156473:	gap of unknown length
*	156474	165327:	contig of 8854 bp in length
*	165328	165427:	gap of unknown length
*	165428	173514:	contig of 8087 bp in length
*	173515	173614:	gap of unknown length
*	173615	180753:	contig of 7139 bp in length
*	180754	180853:	gap of unknown length
*	180854	187919:	contig of 7066 bp in length
*	187920	188019:	gap of unknown length
*	188020	195551:	contig of 7532 bp in length
*	195552	195651:	gap of unknown length
*	195652	202234:	contig of 6583 bp in length
*	202235	202334:	gap of unknown length
*	202335	208703:	contig of 6369 bp in length
*	208704	208803:	gap of unknown length
*	208804	216638:	contig of 7835 bp in length
*	216639	216738:	gap of unknown length
*	216739	221125:	contig of 4387 bp in length

TITLE
JOURNAL

* 10005	11906: contig of 1902 bp in length
* 11907	12006: gap of unknown length
* 12007	14261: contig of 2255 bp in length
* 14262	14361: gap of unknown length
* 14362	16859: contig of 2498 bp in length
* 16860	16959: gap of unknown length
* 16960	19396: contig of 2437 bp in length
* 19397	19496: gap of unknown length
* 19497	22519: contig of 3023 bp in length
* 22520	22619: gap of unknown length
* 22620	24754: contig of 2135 bp in length
* 24755	24854: gap of unknown length
* 24855	30882: contig of 6028 bp in length
* 30883	30982: gap of unknown length
* 30983	35339: contig of 4357 bp in length
* 35340	35439: gap of unknown length
* 35440	41737: contig of 6298 bp in length
* 41738	41837: gap of unknown length
* 41838	45907: contig of 4070 bp in length
* 45908	46007: gap of unknown length
* 46008	54514: contig of 8507 bp in length
* 54515	54614: gap of unknown length
* 54615	60349: contig of 5635 bp in length
* 60250	60349: gap of unknown length
* 60350	69789: contig of 9440 bp in length
* 69790	69889: gap of unknown length
* 69890	79032: contig of 9143 bp in length
* 79033	79132: gap of unknown length
* 79133	91048: contig of 11916 bp in length
* 91049	91148: gap of unknown length
* 91149	102627: contig of 11479 bp in length
* 102628	102727: gap of unknown length
* 102728	116363: contig of 13636 bp in length
* 116364	116463: gap of unknown length
* 116464	131945: contig of 15482 bp in length
* 131946	132045: gap of unknown length
* 132046	155515: contig of 23470 bp in length
* 155516	155615: gap of unknown length
* 155616	219564: contig of 63949 bp in length.
FEATURES	
source	
1. .219564	
/organism="Mus musculus"	
/db_xref="taxon:10090"	
/clone="RP23-187G15"	
/clone_lib="RPci mouse BAC library 23"	
BASE COUNT	55924 a 52052 c 51899 g 56983 t 2706 others
ORIGIN	
Query Match	
Best Local Similarity 31.2%; Score 471.8; DB 2; Length 219564;	
Matches 655; Conservative 0; Mismatches 182; Indels 24; Gaps	
QY 661	ccatgactactgccttggctggcgtgcagaggcaccaggctccttcgcgctggaagc 720
Db 190117	CCACCTCTTCTCTCCTCAGTACGATGCAAGAGGCACCAAGCTCTCTCCGCTGGATGA 190176
QY 721	gcccttgcgcctgtgcgactcgcagagcgccttcctgcctgcctcccttcgc 780
Db 190177	ACCTTCTGTGTCGTGTGTGATTCCTCAAGAACGCCCTGTCTGAGCCTCCCTTTGC 190236
QY 781	cgtcacccgcgcagcctcagccccacctcggcctcattgcctggggaactctcctgggcc 840
Db 190237	CGCACACCGTGCAGCGCTCAGCGCCACCTCAGCATCATCTCTGGCATTTCTTGAGCCC 190296
QY 841	tggtgtgcccaggtgacagtgcacgcctcctctcagccgctcccccagctcgggctc 900
Db 190297	TGCTATTGCCAAGTTCACATGTCACACGCCCTCTCTCAGCCAGTCCCCCAGCTCAGGTTTC 190356
QY 901	tgaggggccagcaggtggacagtggcggggctcagacaccaggagctcccgatggggc 960
Db 190357	TGAGGGTCCAGGCCAGGTAGACAGTGGGAGGGGCTCAGATACTTGAGGGCTCAGAAGGAAT 190416

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2002, 01:06:10 ; Search time 235.79 Seconds
(without alignments)
11002.408 Million cell updates/sec

Title: US-09-786-136-4
Perfect score: 1511
Sequence: 1 tccagatgctgaattctg.....atccatgcaataaagtcca 1511

Indexing table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1511	100.0	1511	21	Human neurotransmi
2	1217.8	80.6	2777	22	Human immune/haema
3	604.8	40.0	864	22	Human neuroblastom
4	370.8	24.5	882	22	Human neuroblastom
5	60.8	4.0	10732	21	Gene encoding a su
6	60.6	4.0	1337	20	Human gene express
7	58.2	3.9	114955	20	Human adenosine A1
8	55.4	3.7	114955	20	Human adenosine A1
9	53	3.5	1127	21	Human colon cancer

c	10	50	3.3	33718	22	AAK65257	Human immune/haema
c	11	50	3.3	33718	22	AAK86411	Human immune/haema
c	12	49.8	3.3	1995	22	AAF63462	Oligonucleotide #2
c	13	49.8	3.3	4686	22	AAK80905	Human immune/haema
c	14	49.8	3.3	4403765	22	AAI99683	Mycobacterium tube
c	15	49.8	3.3	4411529	22	AAI99682	Mycobacterium tube
c	16	48.2	3.2	4684	22	AAK80907	Human immune/haema
c	17	47.2	3.1	2016	23	AAK80907	DNA encoding novel
c	18	46	3.0	15872	18	AAI68715	Streptomyces venez
c	19	46	3.0	15872	21	AAZ87283	S. venezuelae vep
c	20	45.6	3.0	1218	21	AAA02488	Human colon cancer
c	21	45.6	3.0	4403765	22	AAI99683	Mycobacterium tube
c	22	45.4	3.0	989	21	AAK02539	Human colon cancer
c	23	45	3.0	109519	22	AAK08693	Micromonospora DNA
c	24	45	3.0	125401	22	AAI17186	Streptomyces nous
c	25	44.4	2.9	1214	22	AAK60302	Nucleotide sequenc
c	26	44.4	2.9	58857	21	AAK8471	Human zaiapha48 cDN
c	27	44.2	2.9	11320	21	AAZ87298	S. venezuelae macr
c	28	44.2	2.9	36778	21	AAZ87318	S. venezuelae pik
c	29	44.2	2.9	37948	21	AAZ87285	S. venezuelae pik
c	30	44.2	2.9	38506	21	AAZ87285	Nucleotide sequenc
c	31	44.2	2.9	38506	21	AAZ56001	Recombinant cosmids
c	32	44	2.9	1140	21	AAK96228	cDNA encoding a ma
c	33	43.6	2.9	2064	14	AAQ52638	Thermomyces fradi
c	34	43.4	2.9	3600	17	AAI13227	Stromatolite enzym
c	35	43.4	2.9	1359	20	AAZ17254	Brain factor-2
c	36	43.2	2.9	1860	14	AAQ50632	Human gene express
c	37	43.2	2.9	1000	21	AAK04484	Human colon cancer
c	38	43.2	2.9	1352	22	AAI58126	Human polynucleoti
c	39	43.2	2.9	1776	15	AAQ86111	Human alpha-1A adr
c	40	43.2	2.9	1776	16	AAI11601	Human alpha-1A adr
c	41	43	2.8	1474	15	AAQ70311	Mouse NF-ATp. Mus
c	42	43	2.8	1531	21	AAK59261	DNA encoding a hum
c	43	43	2.8	2998	21	AAI21180	Human low adenosin
c	44	43	2.8	2998	21	AAK35058	Human adenosine re
c	45	43	2.8	3478	17	AAK35686	Human transcriptio

ALIGNMENTS

RESULT	1
AAA08405	
ID	AAA08405 standard; cDNA; 1511 BP.
AC	AAA08405;
XX	13-JUL-2000 (first entry)
DE	Human neurotransmitter associated nucleotide sequence SEQ ID NO:4.
XX	Human; neurotransmitter; diagnosis; antiparkinsonian; neuroleptic;
XX	anticonvulsant; nootropic; tranquilizer; neuroprotective; cytostatic;
XX	antidepressant; antidiabetic; gynaecological; immunosuppressive;
XX	neurological disorder; epilepsy; hormone-related disease; Parkinson's disease;
XX	schizophrenia; attention deficit disorder; ss.
XX	attention deficit disorder; ss.
OS	Homo sapiens.
XX	Key
XX	Location/Qualifiers
FT	596..1228
FT	/*tag= a
FT	/product= "neurotransmitter associated protein"
XX	
XX	WO200012685-A2.
XX	
XX	09-MAR-2000.
XX	
XX	26-AUG-1999; 99WO-US19615.
XX	
XX	01-SEP-1998; 98US-0144952.
XX	01-SEP-1998; 98US-0155194.

RESULT 2
AAK84146/c
ID AAK84146 standard; DNA; 2777 BP.
XX
AC AAK84146;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:38958.
XX
KW Human; immune; haematopoietic; Immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX MO200157182-A2.
XX
XX 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225277.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 20-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 01-DEC-2000; 2000US-0233081.
PR 12-SEP-2000; 2000US-02331968.
PR 14-SEP-2000; 2000US-02332397.
PR 14-SEP-2000; 2000US-02332398.
PR 14-SEP-2000; 2000US-02332399.
PR 14-SEP-2000; 2000US-02332400.
PR 14-SEP-2000; 2000US-02332401.
PR 14-SEP-2000; 2000US-02333063.
PR 14-SEP-2000; 2000US-02333064.
PR 14-SEP-2000; 2000US-02333065.
PR 21-SEP-2000; 2000US-02334223.
PR 21-SEP-2000; 2000US-02334274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.

Thu Sep

OS	Vigna angularis.
XX	
PN	JP2000060568-A.
XX	
PD	29-FEB-2000.
XX	
PF	26-AUG-1998; 98JP-0239998.
XX	
PR	26-AUG-1998; 98JP-0239998.
XX	
PA	(WIZU/) MIZUNO K.
XX	
PA	(OJIP) OJI PAPER CO.
XX	
DR	WPI; 2000-342371/30.
DR	P-PSDB; AAY85179.
XX	
XX	A gene encoding a cellulose synthetic equipment - for the improvement
PT	in the amount of cellulose synthesised in a plant body
PT	
PS	Claim 2; Page 14-21; 32pp; Japanese.
XX	
CC	This sequence represents a gene encoding a subunit of the cellulose
CC	synthase complex of Vigna angularis. The invention relates to subunits of
CC	cellulose synthetic equipment, that can be used to increase the amount of
CC	cellulose synthesised by a plant. The proteins and genes encoding them
CC	can also be used to improve the properties of the cellulose being
CC	produced by a plant.
XX	
SQ	sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

5

[illegible]

DT 29-JUN-2000 (first entry)

Cellulose synthase: cellulose production: increase yield; ds.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2002, 00:46:05 : Search time 1848.71 Seconds
(without alignments)
11031.421 Million cell updates/sec

Title: US-09-786-136-4
Perfect score: 1511
Sequence: 1 tccagagtgctgaattctg.....atcctatgcataaagtcca 1511

Indexing table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estlin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: gb_hcc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hcc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	596.6	39.5	841	10	BE798146 601582424
2	560	37.1	561	10	BE793058 1e47c12.y
3	523.8	34.7	527	10	BM353750 1g55f05.y
c	472.8	31.3	495	10	BM127861 1f10c05.x
5	360.8	23.9	600	10	BI990349 4064-13 M
c	319.2	21.1	333	10	BM353476 1g55f05.x
7	317.2	21.0	643	12	BH051641 RPT-24-2
8	311	20.6	494	9	AW246364 2821950.5
c	298.2	19.7	322	9	AW246580 2821950.3
10	294	19.5	321	10	BI792808 1e47c12.x
11	293.6	19.4	649	9	BB307349 BB307349
12	207.8	13.8	526	10	BE080293 213460 MA
13	195.8	13.0	286	10	T75425 yc90g11.r1
14	170.8	11.3	470	10	BF412161 UI-R-BT1-
15	157	10.4	535	10	BE266573 601193567
16	154	10.2	423	10	BF550819 UI-R-C1-j
17	154	9.9	425	9	BB711674 BB711674

18	140	9.3	392	9	BB802514
19	139.8	9.3	348	10	BE862588
20	130.6	8.6	371	9	BB804715
21	130.4	8.6	363	9	BE798951
c	130.4	8.6	410	9	AI044127
22	128.2	8.5	411	9	AI547396
c	124.4	8.2	917	12	CNS045H6
c	120.6	8.0	202	9	BE594108
c	111.1	7.4	387	9	BB800911
26	111	7.3	312	10	BF544431
c	101.4	6.7	406	9	AI797955
28	99.8	6.6	130	10	BF189594
c	94.8	6.3	526	9	AI145213
c	92.2	6.1	316	9	AI851743
30	85	5.6	3760	11	AK016425
32	81.2	5.4	279	9	BB193885
33	79.6	5.3	280	9	BB180183
34	78.2	5.2	311	9	BB317839
c	77	5.1	505	9	AA630176
36	74.2	4.9	323	9	BB182430
37	72.4	4.8	925	12	CNS0091P
c	70.8	4.7	1651	10	BG809816
40	68.2	4.5	289	9	BE270582
c	66.6	4.4	935	12	CNS006XK
42	65.6	4.3	932	12	CNS0072Q
43	65.2	4.3	222	9	BB309157
44	64.6	4.3	221	9	BB193279
45	64.6	4.3	223	4	BB716322

ALIGNMENTS

RESULT 1

BE798146

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE798146 601582424F1 NIH_MGC_7 Homo sapiens CDNA clone IMAGE:3937013 5', linear

BE798146 mRNA sequence.

BE798146 EST.

BE798146.1 GI:10219344

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Carnivora; Homnidae; Homo.

1 (bases 1 to 841)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: DCTD/DP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone Sequencing by: Inceye Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCW781 Row: p Column: 06

High quality sequence start: 38

High quality sequence stop: 795.

Location/Qualifiers

1. .841

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3937013"

/clone_lib="NIH_MGC_7"

/cell_line="MGC3"

/note="DH10B (phage-resistant)"

/host="Organ: Lung; Vector: pORF7; Site:1: XhoI; Site:2:

ECORI; CDNA made by oligo-dt priming. Directionally

cloned into EORI/XhoI sites using the following 5'

adaptor: GGCACGAG(C). Size-selected >500bp for average

FEATURES

source

Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
High quality sequence stop: 446.

FEATURES

Location/Qualifiers
1..495
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 M4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1;
Site 2: Sal 1; Starting library constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
made by oligo-dr priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 95 a 120 c 183 g 97 t

Query Match 31.3%; Score 472.8; DB 10; Length 495;
Best Local Similarity 99.6%; Pred. No. 1.5e-82;
Matches 474; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1035 gatgcgcctgcttcagcagtgaggagtggaactgacagcgcccaaggctgattgctgctgc 1094
DB 495 GATGCCCGCTGCTTCGACGAGTGGGAACACGACAGCGGCCCAAGGCTGATGGCTGGCTGC 436
QY 1095 gggccagcactgctgacggccttgacctggccgctcctcaaggccgagccgagggc 1154
DB 435 GGGCCAGCAGCTGCTGACGGCTTGACCTGGCGGCTCAAGGGCGCAGCCGAGGGC 376
QY 1155 tcttcctgactgagcactgggacaaacactgacgtacactacactgctgtgtacagcc 1214
DB 375 TCTTCTGCTACTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCT 316
QY 1215 cagcagcactgtagaggtgccccctgctgcttgggctggcgcccccccaacacactca 1274
DB 315 CAGCGAACGTGTGAAGGCTGCCCTGCTGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 256
QY 1275 agtcaactgcccagggctgctctggtgctggaagttagctgctgctgctgctgctgctgct 1334
DB 255 AGTCACTGGCGCCAGGGCTGGCCCTCTTGGTGTGGGAAAGTGTAGGCTGGTGGCAGGCT 196
QY 1335 gtccccactgcttcttactcctcctcctggagccctcttggcccccaaaaagtgcctgc 1394
DB 195 GTCCCCACTGCTTCTACTCCTCCTCCTAGAGCCCTCTTGGCCCCCACAAGAGTGCCTGC 136
QY 1395 ctgtctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1454
DB 135 CTGTGCTCTCTCCT 76
QY 1455 gcaacacagtggagggtagagagccacagtcctcccaaatcctatgcaataaagtgc 1510
DB 75 GCAACACAGTGGAGGGTAGAGAGCCACAGTCCCCCAATCTATGCAATAAAGTGC 20

RESULT 5
BI990349

LOCUS

BI990349 600 bp mRNA linear EST 20-DEC-2001
4064-13 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION BI990349
VERSION BI990349.1 GI:17961359
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L., and Klein, W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
COMMENT Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES

source
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
BASE COUNT 111 a 179 c 163 g 130 t 17 others
ORIGIN

Query Match 23.9%; Score 360.8; DB 10; Length 600;
Best Local Similarity 79.4%; Pred. No. 1.2e-60;
Matches 452; Conservative 0; Mismatches 114; Indels 3; Gaps 3;
QY 665 gactaccctgcttggtgctgac-tgacagggacacaggtccttcgctgacagcgc 723
DB 1 GACTACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
QY 724 ctcttgcccgctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 782
DB 61 CTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 783 tgacagcgccagcctgacgcccacccctgctgctgctgctgctgctgctgctgctgctg 842
DB 121 CACACGCTGCACCTGACCCCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 180
QY 843 gtgttgccaggtgacagtgccacgctgctgctgctgctgctgctgctgctgctgctgctg 902
DB 181 NTATTGGCCAAAGGTGACAGTGCCACANCTCTCTGACAGCTGCTGACAGCTGCTGAC 240
QY 903 agggggccagggcaggtggacag-tggggcggggtcagacacccagggctccgaaggcg 961
DB 241 AGGGTCCAGGCCAGGTAGACAGCTGGGAGGGGCTCAGATCTAGAGGCTCATATAGGAATG 300
QY 962 gaagggtcgggcgccagcctgcgggcgccagcctgggcgccagcctgcgtagcactgccc 1021
DB 301 GAAAGCAGGACAGCTCTGATCTTCNNGGGCGCAGTGGGCCACAGNTGNGGNCCTGGCG 360
QY 1022 tggctgagcagcagtgccgctgctgctgctgctgctgctgctgctgctgctgctgctg 1081
DB 361 TGGCTGGAACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 1082 gattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1141
DB 421 GATTGCTGCTGAGAGCCCAACACCTGCCAGGAGCTTGATCTGATCTGCTTCTTAAAGCT 480
QY 1142 gcaagcccgaggctcttctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1201
DB 481 GCTGCCCGGGGGCTCTTCTGTTACTGCGACACTGGGACAGANTCTGACAGCTACACTTG 540

```

QY 1202 ctgtgtacagccagcgaacgtgtgaag 1230
||||| ||||| ||||| ||||| ||||| |||||
Db 541 CTGTGTACACCCCTTCGAATGTGTAAG 569

RESULT 6
LOCUS BM353476/c 333 bp mRNA linear EST 07-JAN-2002
DEFINITION l953f05.x1 HR85 islet Homo sapiens cDNA 3', mRNA sequence.
ACCESSION BM353476
VERSION BM353476.1 GI:18085834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 333)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, G., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 264.
FEATURES
Location/Qualifiers
1..333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HR85 islet"
/tissue_type="Purified pancreatic islet"
/lab_host="pH10B"
Note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
Size: selected on agarose gel. Average insert size -1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."
BASE COUNT 69 a 70 c 130 g 64 t
ORIGIN
Query Match 21.1%; Score 319.2; DB 10; Length 333;
Best Local Similarity 99.1%; Pred. No. 1.4e-52;
Matches 321; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1187 ctgcagctacacctgtgtctacagccagcgaacgtgtgaaggctgccccctgtgct 1246
||||| ||||| ||||| ||||| ||||| |||||
Db 333 CTGCAGCTACACCTGTGTCTACAGCCAGCGACGTGTGAAGGTGCCCTGTGCT 274

QY 1247 tgggtggcccccacccacacactcaagtcaactcgcgcagcagctgacctcttgg 1306
||||| ||||| ||||| ||||| ||||| |||||
Db 273 TGGGCTGGCGCCCGCCACCCACACACTCAAGTCACTGCCGCCACAGACTGGCTCTTGGG 214

QY 1307 ctgggaagtgtaggctgggtgcagcctgtcccccaactcttactccctccctggag 1366
||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 213 CTGGAAAGTGTAGCTGTGCCAGCCTGTCCCCACTGCTTCTTACTCCCTCCTAGAG 154
QY 1367 ccctttgcccccaaaaaagtgcctgtgctgtctctctctctctctctctccacccacac 1436
||||| ||||| ||||| ||||| ||||| |||||
Db 153 CCCTCTTGGCCCCACAAAAAGTGCCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 94
QY 1427 acactccctccatctctgtgagctccctgcacacagtggaagggtagagagccacagtc 1486
||||| ||||| ||||| ||||| ||||| |||||
Db 93 ACACCTCCCTCCATCTCTGTAGCTCCCTGCAACACAGTGGGAAGGTAGAGAGCCACAGTC 34
QY 1487 cccaaatcctatgcaataaagtgc 1510
||||| ||||| ||||| ||||| ||||| |||||
Db 33 CCCAAATCCTATGCATAAAGTGC 10

RESULT 7
LOCUS BH051641 643 bp DNA linear GSS 17-JUL-2001
DEFINITION RPCI-24-260N14.TJ RPCI-24 Mus musculus genomic clone RPCI-24-260N14
, DNA sequence.
ACCESSION BH051641
VERSION BH051641.1 GI:14843393
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 643)
Zhao, S., Geer, K., Krol, M., Shvartsbeyn, A., Akinret, B., Levins, M.,
Tsagayev, G., Nierman, W., Malek, J., Shatsman, S., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.tigr.org/bacpac/orderingframe.html). BAC end
Plate: 260 row: N column: 14
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..643
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-24-260N14"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
Note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 128 a 204 c 178 g 133 t
ORIGIN
Query Match 21.0%; Score 317.2; DB 12; Length 643;
Best Local Similarity 80.3%; Pred. No. 4e-52;
Matches 384; Conservative 0; Mismatches 93; Indels 1; Gaps 1;

QY 661 ccataactacctgcctgtgtgctgagcagagcagcagcagcagcagcagcagcagcagc 720
||||| ||||| ||||| ||||| ||||| |||||

```


Plate: LCM8 row: C column: 7
High quality sequence stop: 183.
Location/Qualifiers
I. .322

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2821950"
/clone_lib="NIH_MGC_7"
/tissue_type="Small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 66 a 67 c 121 g 68 t
ORIGIN

Query Match 19.7%; Score 298.2; DB 9; Length 322;
Best Local Similarity 99.0%; Pred. No. 1.7e-48;
Matches 300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1208 tacagccagcgaacgtgaaagctgcccctgctgctggctggcgcaccccaac 1267
|||||
Db 322 TACAGCCAGCAGCAAGCTGTGAAGCTGCCCTGCTGGCTGGCGCCGCCACCAAC 263
QY 1268 acactcaagtcactgcgcgcagggctgaccttggctgggaaagttaggctggtg 1327
|||||
Db 262 ACATCAAGTCATCTGCCGCCAGGGCTGGCTTGGTGGTGAAGTGTAGGCTGGTT 203
QY 1328 ccagctgtccccactgtcttactccctccctggagccctctgccccacaaaag 1387
|||||
Db 202 CCAGCTGTCCCCACTGCTTCTTACTCCTCCCTAGAGCCCTCTTGGCCCCACAAAAG 143
QY 1388 tgctgctgtgtctctctctctctccacccccactcacactccccctccatctctga 1447
|||||
Db 142 TGGCTGCTGTGCTCTCTCCCTCTCTCCACCCCACTCACATCCCTCCATCCCTTGA 83
QY 1448 gctccctgcacacagtgaaaggttagagccacagtcacaaatccctatgcaataag 1507
|||||
82 GCTCCCTGCACACAGTGGGAAGGTAGAGCCACAGTCCCCCAATCTATGCAATAAG 23
1508 tgc 1510
|||
22 TGC 20

RESULT 10
BI792808/c
LOCUS
DEFINITION
321 bp mRNA linear EST 01-OCT-2001
1e47c12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
CDNA 3', mRNA sequence.
ACCESSION BI792808
VERSION BI792808.1 GI:15820533
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 321)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 286.
Location/Qualifiers
I. .321

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."
BASE COUNT 67 a 69 c 119 g 65 t 1 others
ORIGIN

Query Match 19.5%; Score 294; DB 10; Length 321;
Best Local Similarity 98.0%; Pred. No. 1.1e-47;
Matches 297; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1208 tacagccagcgaacgtgaaagctgcccctgctgctggctggcgcaccccaac 1267
|||||
Db 321 TACAGCCAGCAGCAAGCTGTGAAGGGTGGCCCTGCTGGCTGGCGCCGCCACCAAC 262
QY 1268 acactcaagtcactgcgcgcagggctgaccttggctgggaaagttaggctggtg 1327
|||||
Db 261 ACATCAAGTCATCTGCCGCCAGGGCTGGCTTGGTGGTGAAGTGTAGACTGGTG 202
QY 1328 ccagctgtccccactgtcttactccctccctggagccctcttgcaccccaaaaag 1387
|||||
Db 201 CCAGCTGTCCCCACTGCTTCTTACTCCTCCCTAGAGCCCTCTTGGCCCCACAAAAG 142
QY 1388 tgctgctgtgtctctctctctccacccccactcacactccccctccatctctga 1447
|||||
Db 141 TGCCTGCCTGTGNTTCTCCCTCTCCTCCACCCCACTCACACTCCCTCCATCCTTTGA 82
QY 1448 gctccctgcacacagtgaaaggttagagccacagtcacaaatccctatgcaataag 1507
|||||
Db 81 GTCCTCTGCACACAGTGGGAAGGTAGAGCCACAGTCCCCCAATCTATGCAATAAG 22
QY 1508 tgc 1510
|||
21 TGC 19

Search completed: September 19, 2002, 03:21:01
Job time: 9296 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2002, 00:57:10 ; Search time 54.09 Seconds
(without alignments)
6861.753 Million cell updates/sec

Title: US-09-786-136-4
Perfect score: 1511
Sequence: 1 tcacagatctgaattctg.....atcctatgcaataaagtcca 1511

Mapping table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PTCTUS.COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	3.9	7218	1	US-08-232-463-14
2	49.8	3.3	4403765	4	US-09-103-840A-2
3	49.8	3.3	4411529	4	US-09-103-840A-1
4	46	3.0	15872	4	US-09-105-537-1
5	45.6	3.0	4403765	4	US-09-103-840A-2
6	44.2	2.9	11220	4	US-09-105-537-32
7	44.2	2.9	36778	4	US-09-105-537-5
8	44.2	2.9	38506	3	US-09-320-878-19
9	44	2.9	2064	1	US-08-343-428-1
10	43.6	2.9	3600	1	US-08-537-002A-5
11	43.6	2.9	3600	3	US-08-863-010-5
12	43.6	2.9	3600	4	US-09-024-429-5
13	43.4	2.9	1860	2	US-08-331-644-3
14	43.4	2.9	1860	5	PCR-US93-04102-3
15	43.2	2.9	1776	1	US-08-722-001-29
16	43	2.8	1288	1	US-08-440-856A-9
17	43	2.8	1474	5	PCR-US94-00545-19
18	43	2.8	3478	1	US-08-396-479B-1
19	43	2.8	3478	1	US-08-818-823-1
20	42.8	2.8	2823	1	US-08-398-008A-1
21	42.8	2.8	2823	1	US-08-893-333-1
22	42.8	2.8	4411529	4	US-09-103-840A-1
23	42	2.8	2889	1	US-08-537-002A-4
24	42	2.8	2889	3	US-08-863-010-4
25	42	2.8	2889	4	US-09-024-429-4
26	41.8	2.8	1162	2	US-08-726-306A-52
27	41.8	2.8	1480	2	US-08-839-008-8

28	41.8	2.8	2923	1	US-08-377-292-6
29	41.8	2.8	2923	2	US-07-989-847-7
30	41.8	2.8	2923	4	US-08-469-411-7
31	41.4	2.7	3382	2	US-08-682-847-1
32	41.4	2.7	43280	2	US-08-804-227C-1
33	41.2	2.7	780	4	US-09-323-555B-1
34	41.2	2.7	53526	3	US-08-658-136-2
35	41.2	2.7	53577	3	US-08-658-136-1
36	41	2.7	44377	2	US-08-804-227C-7
37	41	2.7	44377	2	US-08-804-198-1
38	40.6	2.7	49136	4	US-09-422-869-1
39	40.2	2.7	1537	2	US-08-839-008-1
40	40.2	2.7	2923	6	5187076-5
41	40	2.6	1288	1	US-08-440-856A-9
42	40	2.6	1402	1	US-08-447-965A-1
43	40	2.6	1417	4	US-09-199-737-3
44	39.8	2.6	1227	3	US-09-074-912-3
45	39.8	2.6	1227	4	US-09-290-136-3

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14

[illegible]

```

RESULT      3
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1

```

```

; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      3.3%; Score 49.8; DB 4; Length 4411529;
Best Local Similarity 48.2%; Pred. No. 0.089;
Matches 172; Conservative 0; Mismatches 182; Indels 3; Gaps 1

QY    695   gcaacaggctccttcgcgtggaacgcgcccttctgcgcgcgtgtgcgcattccgcaggag 754
      |||||
Db    337017 gccgcggcgccgaacagcccgcgtgcgcccccgcgcgcgccttcgggcgcgagcgtgcc 337076

QY    755   cgcctctgcgtgcctcgcacctttgcgtgcaccgcgcagcactcagccccacctcggcc 814

```

[illegible]

[illegible]

RESULT

```

US-08-343-428-1
Sequence 1, Application US/08343428
Patent No. 5665586
GENERAL INFORMATION:
APPLICANT: Nakamura, Etsuo
APPLICANT: Tsuzuki, Hiroshige
APPLICANT: Kitadokoro, Kengo
APPLICANT: Shin, Nasaru
APPLICANT: Teraoka, Hiroshi
TITLE OF INVENTION: No. 5665586el Protease
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: James F. Haley, Jr., Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS Dos 5.0
SOFTWARE: Wordperfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,428
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/J93/00592
FILING DATE: 30-APR-1993
APPLICATION NUMBER: JAPAN 4-126511
FILING DATE: 19-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Haley, Jr., James F.
REGISTRATION NUMBER: 27794
REFERENCE/DOCKET NUMBER: SHGN-7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)596-9000
TELEFAX: (212)596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2064
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:

```


OM of: US-09-786-136-4 to: SPTREMBL_19:* out_format : pfs
Date: Sep 19, 2002 5:45 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=fragnet_n2p.model -DEV=xlh
-Q/cgn2_1/USPRO_spool/US09786136/runat_16092002_141244_19221/app_query.fasta_1.1592
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DEL0P=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-MAX=100 -THR_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs
-RM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09786136 -CGN1_1.100 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-786-136-4
Query length: 1511
Database: SPTREMBL_19*
Database length: 567222
Search time (sec): 94.020000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_human:Q9BVH8	+ 1778.50	2027.29	1.4e-105	1001	! Q9BVH8 homo sapiens (human).
sp_mammal:Q9GWT9	+ 1625.50	1854.11	7.7e-96	803	! Q9GWT9 macaca fascicularis (cre
sp_mammal:Q9GKV4	+ 1277.00	1464.88	1.1e-73	260	! Q9GKV4 macaca fascicularis (cre
sp_rodent:Q9CUE8	+ 359.50	402.55	3.7e-15	1209	! Q9CUE8 mus musculus (mouse).
sp_human:Q9BX42	+ 240.50	271.89	1.3e-07	628	! Q9BX42 homo sapiens (human).
sp_plant:Q9PQ6	+ 235.00	266.63	3.0e-07	555	! Q9PQ6 chlamydomonas reinhardtii
sp_bacteria:Q9DXH2	+ 233.00	261.68	4.1e-07	763	! Q9DXH2 mycobacterium tuberculosis
sp_plant:Q9SPM0	+ 226.50	249.71	1.1e-06	1315	! Q9SPM0 zea mays (maize). exten
sp_fungi:Q41935	+ 224.00	253.61	1.5e-06	585	! Q41935 murid herpesvirus 4. hyp
sp_fungi:Q42854	+ 218.00	238.29	3.9e-06	1611	! Q42854 schizosaccharomyces pom
sp_virus:P89459	+ 218.00	232.77	4.1e-06	3122	! P89459 herpes simplex virus (u
sp_bacteria:Q68872	+ 217.50	246.81	3.9e-06	542	! Q68872 myxococcus xanthus. hyp
sp_invertebrate:Q9VZC2	+ 215.00	245.57	5.5e-06	446	! Q9VZC2 drosophila melanogaste
sp_invertebrate:Q9VPG1	+ 215.00	231.95	6.2e-06	2284	! Q9VPG1 drosophila melanogaste
sp_human:Q9H6K5	+ 214.00	241.74	6.5e-06	616	! Q9H6K5 homo sapiens (human). co
sp_invertebrate:Q57580	+ 210.50	238.32	1.1e-05	574	! Q57580 gallus gallus (chicken)
sp_virus:Q65553	+ 210.00	223.29	1.3e-05	3247	! Q65553 bovine herpesvirus 1. u
sp_plant:Q41805	+ 209.50	231.11	1.3e-05	1188	! Q41805 zea mays (maize). exten
sp_virus:Q41972	+ 207.00	232.59	1.8e-05	706	! Q41972 murid herpesvirus 4. hyp
sp_virus:Q41973	+ 207.00	232.35	1.8e-05	727	! Q41973 murid herpesvirus 4. hyp
sp_virus:Q41971	+ 207.00	231.82	1.8e-05	774	! Q41971 murid herpesvirus 4. hyp
sp_virus:Q69146	+ 204.00	223.16	2.8e-05	437	! Q69146 human herpesvirus 4 (eps
sp_mammal:Q95JDI	+ 203.00	229.86	3.3e-05	566	! Q95JDI sus scrofa (pig). basic
sp_human:Q75851	+ 202.00	212.15	4.3e-05	4123	! Q75851 homo sapiens (human). w
sp_invertebrate:Q9B177	+ 201.50	217.81	4.4e-05	1953	! Q9B177 nephila madagascari
sp_virus:Q905L3	+ 201.00	227.02	4.4e-05	605	! Q905L3 cercopithecine herpesvir
sp_virus:Q91PQ9	+ 201.00	226.83	4.4e-05	619	! Q91PQ9 cynomolgus epstein-barr
sp_mammal:Q95JDI	+ 199.50	226.71	5.4e-05	511	! Q95JDI sus scrofa (pig). basic
sp_invertebrate:Q93107	+ 197.50	221.19	7.4e-05	753	! Q93107 acanthamoeba castellan
sp_bacteriophage:Q9RX57	+ 197.50	220.29	7.5e-05	839	! Q9RX57 deinococcus radiodurans
sp_human:Q9U0T1	+ 197.50	217.85	7.6e-05	1124	! Q9U0T1 homo sapiens (human). k
sp_mammal:Q95JDI	+ 196.50	220.94	8.5e-05	676	! Q95JDI sus scrofa (pig). basic
sp_invertebrate:Q9N968	+ 196.00	218.79	9.3e-05	817	! Q9N968 leishmania major. hyp
sp_human:Q9UDT9	+ 195.50	216.08	0.0001	1056	! Q9UDT9 homo sapiens (human). w
sp_bacteria:Q9F2N5	+ 194.00	215.80	0.0001	899	! Q9F2N5 streptomyces coelicolor
sp_human:Q9V5L9	+ 193.50	205.16	0.0001	2871	! Q9V5L9 homo sapiens (human). t
sp_human:Q9F206	+ 193.00	213.31	0.0001	1044	! Q9F206 homo sapiens (human). k
sp_rodent:Q9R172	+ 193.00	206.66	0.0002	2319	! Q9R172 rattus norvegicus (rat)
sp_fungi:Q9P6T1	+ 192.50	207.35	0.0002	1992	! Q9P6T1 neurospora crassa. hyp

sp_plant:Q9FLO7 + 192.00 210.41 0.0002 1289 ! Q9FLO7 arabidopsis thaliana
sp_invertebrate:Q9NE65 + 191.50 207.16 0.0002 1778 ! Q9NE65 leishmania major.
sp_rodent:Q63123 - 191.00 208.47 0.0002 1419 ! Q63123 rattus norvegicus (r
sp_invertebrate:Q9BIU7 - 189.50 213.29 0.0002 648 ! Q9BIU7 argiope trifasciata
sp_rodent:P70433 + 189.50 211.51 0.0002 802 ! P70433 mus musculus (mouse).
seq_name: sp_human:Q9BVH8
seq_documentation_block:
ID Q9BVH8 PRELIMINARY; PRT; 1001 AA.
AC Q9BVH8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEICAL 106.2 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE, RETINOBLASTOMA;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL:BC001200; AAC01200.1;
DR InterPro:IPR002035; WFA.
DR SMART:SM00327; WFA; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1001 AA; 106194 MW; 9DFD75263D040E43 CRC64;

alignment_scores:
Quality: 1778.50 Length: 400
Ratio: 5.111 Gaps: 3
Percent Similarity: 87.000 Percent Identity: 85.250
alignment_block:
US-09-786-136-4 x Q9BVH8
Align seg 1/1 to: Q9BVH8 from: 1 to: 1001
86 TGGGGAGCTGAGCTGGTCTGGGGCTAGCAGGCAATCTACTCTGTTAA 135
||||| ||||| ||| :
623 TrpGlyProGlyAspGlySerGlnProProSer..... 633
136 AGTCGATAGAGGAGGAGAGCTGGGGAGACACTGCAATTTTCTATGAGC 185
634 PROP 635
186 CCTCTGTAGAGGTGGTGG...TGGGGGCTCCAGACTGCTCTGGT 232
|| |||:||||| :
635 roValArgGluAlaTrpAspGlnAlaLeuHisArgLeuThrAlaAla 651
233 TCCACA..... 238
652 ServaValArgAspAsnGluGlnLeuAlaLeuArgGlyAlaGluTh 668
239GGCCATGCCCCGAGGTGCTGGCTTCAGGCCCTTCAGA 275
668 rThrAlaAspArgGlyHisAlaArgArgCysTrpLeuArgAlaLeuGlnT 685
276 CAAGTAAGTCAAGCTGCTGCCCTCTCTGCTTCACTTGCCTGTAGCTGTG 325
685 hrSerLysValSerAlaProSerCysPheThrCysProValAlaVal 701
326 GATGCTACTACTAGGAGGTCTGCTGGGGCCCTGCAGGTGTCAGCTC 375
702 AspAlaThrThrArgGluValLeuProGlyAlaLeuGlnValCysSerSe 718
376 AGAGCCCGCTGAGCCCGCCAGGAAACCCCTCTCTGCTCTCACGCCATCTAG 425
718 rGluProAlaGluProProGlyThrProProAlaSerHisSerHisLeuA 735

```

426 ATGCAGCTCTCTGCCCACTGTGTGTACTTAAGGACTTCAGAGAGC 475
|||||
735 spAlaAlaProLeuProThrValValTyrSerLysGlyLeuGlnArgGly 751
|||||
476 TCTCCAGCAGCGCTCGGACTCGACCAACAAATGCCAACTCCAGCGTGC 525
|||||
752 SerProAlaGlyAlaTrpAspSerAspGlnAsnGlyAsnSerLysArgAl 768
|||||
526 TTTGGGGACCTCGCACTCCACCAAGAGTCTCGCCGCCCACTCCGCC 575
|||||
768 aLeuGlyAspProAlaThrProThrGluGlyProArgArgProProAla 785
|||||
576 GTCTCTCTCTCGGCTCAGCATGCGCGCCGCTCACAACACTCTGAGCCCT 625
|||||
785 rgProProCysArgLeuSerMetGlyArgGHisLysLeuCysSerPro 801
|||||
626 GACCCGGCCAGCCCAACACAGTCAAGCCAGCGACCATGACTACCTGCC 675
|||||
802 AspProGlyGlnAlaAsnSerGluGlySerAspHisAspTyrLeuPr 818
|||||
676 CTTGGTGGCGCTGCGAGGAGCACCAGGCTCCTTCGCCCTGACGCGCCCT 725
|||||
818 oLeuValargLeuGlnGluAlaProGlySerPheargLeuaspAlaProp 835
|||||
726 TCTGCGCGCTGTGGCATCTCGCAGGAGCGCTCTGCCGTGCGCTCGCCC 775
|||||
835 heCysAlaAlaValArgTleSerGlnGluArgLeuCysArgAlaSerPro 851
|||||
776 TTTGGCGTGCACCGCGCAGCTCAGCCCACTCGGCTCATTTGCCCTG 825
|||||
852 PheAlaValHisArgAlaSerLeuSerProThrSerAlaSerLeuProTr 868
|||||
826 GGCACCTTCTGGGCCCTGTGTGGCAGGGTGCACAGTCCACGCGCTCCT 875
|||||
868 paLaLeuLeuGlyProGlyValGlyGlnGlyAspSerAlaThrAlaSerC 885
|||||
876 GCAGCCGCTCCCGCAGCTCGGGCTCTGAGGGCCAGCCAGCGTGGACAGT 925
|||||
885 ysSerProSerProSerGlySerGlyGlyProGlyGlnValAspSer 901
|||||
926 GGGCGGGCTCAGACACCGAGCGCTCCGATGGGCGGAAGCGCTGGGCGG 975
|||||
902 GlyArgGlySerAspThrGluAlaSerGluGlyAlaGluGlyLeuGlyG 918
|||||
976 CACCGACCTGGCGGCGCGGACCTGGGCGACCTGCCGTAGCAGCTGCCGTGC 1025
|||||
918 yThrAspLeuArgGlyArgThrTrpAlaThrAlaValAlaLeuAlaTrpL 935
|||||
1026 TGGAGCAGCGATGCCCGCTGCCCTCGACGAGTGGGAAGTGCAGCGGCC 1075
|||||
935 euGluHisArgCysAlaAlaAlaPheAspGluTrpGluLeuThrAlaAla 951
|||||
076 AAGGCTGATTGCTGCTGGCGGCCAGCACCTTGCTGACGCGCTTGACCT 1125
|||||
952 LysAlaAspCysTyrPleuArgAlaGlnHisLeuProaspGlyLeuaspLe 968
|||||
1126 GCGCCCGCTCAAGCGCGCGCGCGGAGGCTCTTCTGCTACTGGCGCCACT 1175
|||||
968 uAlaAlaLeuLysAlaAlaAlaArgGlyLeuPheLeuLeuLeuArgHisT 985
|||||
1176 GGGACCAAAACCTGAGCTACACCTGCTGCTGCTACGCGCCAGCGAAGTG 1225
|||||
985 rpAspGlnAsnLeuGlnLeuHisLeuLeuCysTyrSerProAlaAsnVal 1001
|||||

```

seq_name: sp_mammal:Q9GMT9

seq_documentation_block:

AC Q9GMT9; PRELIMINARY;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

PRT; 803 AA.

```

DE HYPOTHETICAL 84.7 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_Taxid=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM CORTEX;
RA Osaka N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RL libraries."
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB047829; BAB12255.1; -
KW Hypothetical protein.
SQ SEQUENCE 803 AA; 84655 MW; 724FC68DCC2D1D92 CRC64;

```

alignment_scores:

Quality: 1625.50

Ratio: 4.926

Percent Similarity: 82.500

Percent Identity: 79.500

alignment_block:

US-09-786-136-4 x Q9GMT9 ..

Align seg 1/1 to: Q9GMT9 from: 1 to: 803

```

86 TGGGGAGTCTGAGCTGTTCTGGCGCTAGCAGGCAATCTACTCTGTAA 135
|||||
433 TrpGlyProGlyAspGlySerGlnProSer..... 443
|||||
136 AGTCGATAGAGGAGAGCTGGGAGAACACTGCAATTTTCTATGAGC 195
|||||
444 .....Prop 445
|||||
186 CCTCTGTAGAGGTGGGTG...TGGGGGCTCCACAGACTGCTCTTGGT 232
|||||
445 roValArgGluAlaAlaTrpAspGlnAlaLeuHisArgLeuThrAlaAla 461
|||||
233 TCCACA..... 238
|||||
462 SerValValArgAspAsnGluGlnLeuAlaLeuArgGlyGlyAlaGluTh 478
|||||
239 .....GGCATCCCGGAGGTGCTGGCTTCGAGCCCTTCAGA 275
|||||
478 rThrAlaAspArgGlyHisAlaArgCysTrpLeuArgAlaLeuGlnT 495
|||||
276 CAAGTAAGTTCAGCTCTGCCCTCTCTGCTTCTACTTGCCTGTAGTGTG 335
|||||
495 hrSerLysValSerSerAlaProSerCysPheThrCysProValAlaVal 511
|||||
326 GATGCTACTACTAGGAGGTCTGCTGGGCGCTCGAGGTGTGCAGCTC 375
|||||
512 AspAlaThrThrArgGluValLeuProGlyValLeuGlnValCysSerSe 538
|||||
376 AGAGCCCGCTGAGCCCGCAGAACCCCTCTGCTCTCACAGCCATCTAG 425
|||||
528 xGluProAlaGluProGlyThrProProAlaAlaHisSerArgLeuA 545
|||||
426 ATGAGCTCTCTGCGCACCTGTGTGTCTACTCTTAAGGACTTTCAGAGGCG 475
|||||
545 spAlaAlaProLeuProThrValValTyrSerLys..... 556
|||||
476 TCTCCAGCAGCGCTGGGACTCGGACCAAAATGGCAACTCCCAAGCGTGC 525
|||||
557 .....GlyAlaTrpAspSerAspArgAsnGlyAsnSerLysCysAl 570
|||||
526 TTTGGGGACCGCTGCCACTCCCAAGAGTCTCGCGCCCGCCACCTCCCC 575
|||||
570 aLeuGlyAspAlaAlaThrProMetGluGlyProArgCysProProProA 597
|||||

```



```

1078 ArgThrArgValSerSerProAlaPro...ValValLysCysCysProPr 1093
      |||      :|||      :|||      :|||      :|||      :|||
1100 CAGCATTGCTGAGCGGCTGACCTGACCCGCTGACGAGCCGAGCCG 1149
      |||      :|||      :|||      :|||      :|||      :|||
1093 opProHleuValSerSer.....ProPro.....ProAlaPro 1104
1150 AGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1176
      :|||      :|||      :|||      :|||      :|||      :|||
1105 LysSerLeuProProHleuProHleuProHleuProHleuProHleu 1121
1177 .....GGACCAAA. 1185
1121 sSerSerProProProHleuProHleuProHleuProHleuProHleu 1138
1186 .....CCTGAGCTACACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1230
      |||      :|||      :|||      :|||      :|||      :|||
1138 eSerSerProProProHleuProHleuProHleuProHleuProHleu 1154
1231 GCTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1280
      :|||      :|||      :|||      :|||      :|||      :|||
1155 Ser.ProProProHleuProHleuProHleuProHleuProHleuPro 1171
1281 TGCGGCCAGAGGCTGCGCTC..TTGCTGCTGGAAGGTAGCTGCTG 1327
      |||      :|||      :|||      :|||      :|||      :|||
1171 erProProProHleuProHleuProHleuProHleuProHleuProHleu 1187
1328 CCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1365
      |||      :|||      :|||      :|||      :|||      :|||
1188 ProProProHleuProHleuProHleuProHleuProHleuProHleu 1204
1366 GCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1412
      |||      :|||      :|||      :|||      :|||      :|||
1204 opProHleuProHleuProHleuProHleuProHleuProHleuPro 1221
1413 CTGCCACCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1439
      |||      :|||      :|||      :|||      :|||      :|||
1221 roProHleuProHleuProHleuProHleuProHleuProHleuPro 1237
1440 .....TCCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
      |||      :|||      :|||      :|||      :|||      :|||
1237 opProHleuProHleuProHleuProHleuProHleuProHleuPro 1251
1477 AGCCACAGTCCCAATCT 1496
      |||      :|||      :|||      :|||      :|||      :|||
1251 roProSerLeuProProPro 1257
seq_name: sp_virus:041935
seq_documentation_block:
ID 041935 PRELIMINARY; PRT; 585 AA.
AC 041935;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOHETICAL 60.2 KDA PROTEIN.
GN GAMMAHV.M6.
OS murid herpesvirus 4.
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=33708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WDMS;
RX MEDLINE=9736649; PubMed=9223479;
RA Virgin H.W. IV, Latreille P., Mamsley P., Hallsworth K., Weck K.E.,
  Dal Canto A.J., Speck S.H.,
  *Complete sequence and genomic analysis of murine gammaherpesvirus
  68.4.
RT J. Virol. 71:5894-5904 (1997).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN-WDMS;
RC

```

RA Latreille P., Mamsley P., Waterston R.H.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U97553; AAB6392.1; -;
 KW Hypothetical protein.
 SO SEQUENCE 585 AA; 60160 MW; 85610A80C34827D CRC64;

alignment_scores:
 Quality: 224.00 Length: 498
 Ratio: 1.093 Gaps: 32
 Percent Similarity: 41.165 Percent Identity: 27.309

alignment_block:
 US-09-786-136-4 x 041935 ..

Align seg 1/1 to: 041935 from: 1 to: 585

```

238 AGGCAATGCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 287
      |||      :|||      :|||      :|||      :|||      :|||
141 ArgProProProHleuProHleuProHleuProHleuProHleuProHleu 154
288 GCTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 337
      |||      :|||      :|||      :|||      :|||      :|||
155 .....AlaProSerArgAlaGlyAlaArgIle 163
338 AGGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 387
      :|||      :|||      :|||      :|||      :|||      :|||
164 ProAsp...LeuProHleuProHleuProHleuProHleuProHleuProHleu 178
388 GCCCCAGAGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437
      |||      :|||      :|||      :|||      :|||      :|||
178 gProProHleuProHleuProHleuProHleuProHleuProHleuProHleu 195
438 TGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 487
      |||      :|||      :|||      :|||      :|||      :|||
195 laPro.....SerArgAlaGly 200
488 GCCTGGAGCTCGAGCCAAATGGCACTCCAGCGCTTGGGGAGCC 537
      |||      :|||      :|||      :|||      :|||      :|||
201 AlaArgIleProHleuProHleuProHleuProHleuProHleuProHleu 208
538 TGCCATCTCCAGGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 587
      :|||      :|||      :|||      :|||      :|||      :|||
209 .ProLeuProSerTrpIleProHleuProHleuProHleuProHleuProHleu 225
588 GGCTGACATGGGCGCCGCGACCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 637
      :|||      :|||      :|||      :|||      :|||      :|||
225 IuLeuGlyProGly.....SerProHleuProHleuProHleuProHleu 238
638 GCCACACAGTGAGAGGAGCGACGACATGACTGCTGCTGCTGCTGCTGCTGCTG 687
      |||      :|||      :|||      :|||      :|||      :|||
239 AlaGly.....AlaArgIle 243
688 GCAGGAGGACACGAGCTCT..TCCGCTG..... 716
      :|||      :|||      :|||      :|||      :|||      :|||
243 eProAspLeuProGly..ProLeuProSerTrpIleProHleuProHleuPro 259
717 .....ACGGCGCTTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 748
260 ProArgProProHleuProHleuProHleuProHleuProHleuProHleuPro 276
749 CAGAGAGCGCTGCTGCT.....GTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
      :|||      :|||      :|||      :|||      :|||      :|||
276 oSerArgAlaGlyAlaArgIleProHleuProHleuProHleuProHleuProHleu 293
787 CCGCGCAGAGCTGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 836
      :|||      :|||      :|||      :|||      :|||      :|||
293 rPleuProAspProArgProArgProArgProArgProArgProArgProArgPro 307
837 GCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 886
      |||      :|||      :|||      :|||      :|||      :|||
307 lyProGly.....SerProHleuProHleuProHleuProHleuProHleuProHleu 315

```

```

887 CCAGCTGGGCTCT.....GAGGGCCAGCCAGGTGACAG 924
      ||||| : : : : : ||||| : : ||
316 ProSerAlaGlaGlyAlaArgIleProAspLeuProGlyProLeuProSe 332
      ||||| : : : : : ||||| : : ||
925 TGGGGGGGGCTCAGACCGAGCGCTCCGANTG.....GGGGGAGAGGC 968
      ||||| : : : : : ||||| : : ||
332 Trp...GlyProAsp.ProArgProProArgProProProLeuGly 347
      ||||| : : : : : ||||| : : ||
969 TGGGGGGGACCGACTGGGGGGCGGA..... 995
      ||||| : : : : : ||||| : : ||
348 ProGlySerProThrSerProAlaProSerAlaGlaGlyAlaArgIleR 364
      ||||| : : : : : ||||| : : ||
996 .....CTGGGGCCACTGCCCTAGCAGCTCAGCTGAGGACAGCGAG. 1037
      ||||| : : : : : ||||| : : ||
364 AspLeuProGlyProLeu.....ProSerTrpGlyProAsp 377
      ||||| : : : : : ||||| : : ||
1037 ..... 1037
377 roArgProProArgProProProGlyLeuGlyProGlySerProThrSer 393
      ||||| : : : : : ||||| : : ||
1038 .....GCCGCCGT 1045
394 ProAlaProSerAlaGlaGlyAlaArgIleProAspLeuProGlyProLe 410
      ||||| : : : : : ||||| : : ||
1046 GCCTTCGACGAGTGGGAGACTGACAGCGCCAGGCTGATTGCTGGCTGGC 1095
      ||||| : : : : : ||||| : : ||
410 uProSerTrpGlyProAspProArgProProArgProProProGlyLeuG 427
      ||||| : : : : : ||||| : : ||
1096 GGGCCAGCACTTGCTGACGGCGCTTGGCTGAGCGCCCTCAAGGCGCGAG 1145
      ||||| : : : : : ||||| : : ||
427 LyProGlySerProThrSer.ProAlaProSerAlaGlaGlyAlaArgI 443
      ||||| : : : : : ||||| : : ||
1146 CCCGAGG.....GCTCTTCTCTGCTACTGCGCCACTGG 1177
      ||||| : : : : : ||||| : : ||
443 eProArgProSerAlaGlyProProProGlyLeuGlyProGlySerProA 460
      ||||| : : : : : ||||| : : ||
1178 GACCAAAACATGACCTACACCTGCTGCTGCTACAGCC...AGCGAGCT 1224
      ||||| : : : : : ||||| : : ||
460 heProArgProSerAlaGlyProProProGlyLeuGlyProGlySerPro 476
      ||||| : : : : : ||||| : : ||
1225 GTGAAGGCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1274
      ||||| : : : : : ||||| : : ||
477 LeuProGly.ProLeuProSerTrpGlyProAspProProThrPhe... 491
      ||||| : : : : : ||||| : : ||
1275 AGTCACTGCCCGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1324
      ||||| : : : : : ||||| : : ||
492 ..ProAlaProSerAlaGly.....AlaArg 500
      ||||| : : : : : ||||| : : ||
1325 GTGCCAGCTGTCCCGCCAGCTGCTTCTACTCCCTCCCTGAGCCCTCTG 1374
      ||||| : : : : : ||||| : : ||
501 IlePro.ProValProProThrPhe.....ProA 510
      ||||| : : : : : ||||| : : ||
1375 CCCCCCAAAAAGTCCGCTGCTGCTGCTCTCTCCCTCCCTCCAGCCAG 1424
      ||||| : : : : : ||||| : : ||
510 laProSerAlaGlaGlyAlaArgIleProArgProProArgProProPro 526
      ||||| : : : : : ||||| : : ||
1425 TCACAGCTCCCTCCATCTCTGAGTCCCTGCA.....ACACATGGAA 1468
      ||||| : : : : : ||||| : : ||
527 GluLeuGlyProGlySerProAspLeuProGlyProLeuProSerTrpG 543
      ||||| : : : : : ||||| : : ||
1469 GCGTAGACAGCCAGAGTCC...CCAATCTCT 1496
      ||||| : : : : : ||||| : : ||
543 yProValProProProSerAlaArgProSerPro 553
      ||||| : : : : : ||||| : : ||

seq_name: sp_fung1:042854

```

```

GN SPAC23A1.17.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCB1_TaxID=4896;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021813; CAA16991.1;
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3.1.
DR PRINTS: PRO0452; SH3DOMAIN.
DR SMART: SM00326; SH3.1.
DR PROSITE: PS50002; SH3.1.
KW Hypothetical protein.
SQ SEQUENCE 1611 AA; 170526 MW; B3DEF9F83C1A7542 CRC64;

```

```

alignment_scores:
  Quality: 218.00      Length: 460
  Ratio: 0.991        Gaps: 26
  Percent Similarity: 47.826      Percent Identity: 26.957

```

alignment_block:
US-09-786-136-4 x 042854 ..

Align seg 1/1 to: 042854 from: 1 to: 1611

```

314 CCTGTAGCT..GTGGATGCTACTACTAGGAGAGTCTGCTGGGGCCT 360
      ||||| : : : : : ||||| : : ||
812 ProValSerIleValThrSerGlyArgProAlaLeuProAspLeu 828
      ||||| : : : : : ||||| : : ||
361 GCAGGTGTCAGCTCAGAGCCGCTGAGCC...CCAGAACCCCTCTCG 407
      ||||| : : : : : ||||| : : ||
828 TalAspProSerSerSerIleGlyHisProLeuProSerProProPro 845
      ||||| : : : : : ||||| : : ||
408 CCTCT.....CACAGCCATTA 424
      ||||| : : : : : ||||| : : ||
845 laAspPheAsnSerIleuAsnValaLaspPheTyrGluProHisSerTyr 861
      ||||| : : : : : ||||| : : ||
425 GATGCA...GCTCCTCTGCC..... 442
      ||||| : : : : : ||||| : : ||
862 GluSerProAlaProGluProGluProSerTyrGluGluGluSerPhe 878
      ||||| : : : : : ||||| : : ||
443 ...ACTGTTGCTACTCTAAAGACTTCAGAGAGCTCTCCAGCAGGG 488
      ||||| : : : : : ||||| : : ||
878 malathrValIleHisAlaProThrProSer.ThrAlaThrPheGlnGly 894
      ||||| : : : : : ||||| : : ||
489 CCTGGGACTCGGACCAAAATGGCACTCCAGCGTGTGGGGGACCT 538
      ||||| : : : : : ||||| : : ||
895 HAspProThrIleSerAsnValaAlaThrPro..... 904
      ||||| : : : : : ||||| : : ||
539 GCCATCCCAAGGAGGTCTCGCGCCCACTCCCGCTCTCCCTGTCG 588
      ||||| : : : : : ||||| : : ||
905 ..ProLeuGlyGlnAspValThrGluSer.....LysAlaSerProVal 919
      ||||| : : : : : ||||| : : ||
589 GCTCAGCATGGGCGCGCTCACAAACTGTGTAGCCCTGACCGG..... 632
      ||||| : : : : : ||||| : : ||
919 laAspAlaSerAlaThrHisGlnSerSerThrGlyLeuThrGlnGluIle 935
      ||||| : : : : : ||||| : : ||
633 ...GCCAGGCAACAACAGTGAAGGAGGAGGACCATGACTACTGCTGCTG 679
      ||||| : : : : : ||||| : : ||
936 ThrGlnLeuGlySerAsnMetArgLeuProThrHisLeuThrArgProSe 952
      ||||| : : : : : ||||| : : ||
680 GTGGCGCTGCAGAGAGCAGAGCTCTCTCCGCTGAGAGCCGCTTCTG 729
      ||||| : : : : : ||||| : : ||
952 rAsnAspGlyArgLysAlaSerGlyPro.....ArgProAla 965
      ||||| : : : : : ||||| : : ||
730 GCGCCGCTGTGCGGCATCTGCGAGAGGCGCTCTGCGGCTGCTGCGCTT 779
      ||||| : : : : : ||||| : : ||

```

```
965 laPro.....ProSerlleProProleu 973
780 CCG.....TGCACCGCGCAGCCTCACCCACCTCGGCTCA 817
    |||
    |||
974 ProValSerAsnIleuSerSerProThrSerIuProProlysAspHl 990
818 TTGGCCTGGGCACTTGGGCGCTGTGGCCAGGGGTGACAGTGCAC 867
    |||
    |||
990 sProProSerAla.....ProleuSerLysProValSerThrSerProA 1005
868 GGCCTCCGCGAGCCGT...CCCCAGCTCGGCTCGAGGGGCCAGGCC 914
    |||
    |||
1005 laAlaProleuAlaIarGValProProValProlYsleuSerSerLysAla 1021
915 AGTGTGACAGTGGCGGGGCTCAGACACGAGCCCTCGATGGGCGGAA 964
    |||
    |||
1022 .....ProProVal..... 1024
965 GGGCTGGGGCGGACCGACCTGCGGGCGGCGGACCTGGGCGCTAGC 1014
    |||
    |||
1025 .....ProleuProSerAlaAspAlaProProIleProValPro.... 1037
1015 ACTGGCGCTGGGAGCAGCGATGCGCGGCTCGCTCGAGAGTGGGAGC 1064
    |||
    |||
1038 .....SerThrAlaProProValProIleProThrSerThr 1049
1065 TGACAGCGCGCCAGCGTGTGCTGCTGCTGC..... 1094
    |||
    |||
1050 ProProValProlYsSerSerSerGlyAlaProSerAlaProProProVa 1066
1095 ...GGGCGCCAGCACTGCTGACGCGCTGACCTGGCGGCGCTCAAGGC 1140
    |||
    |||
1066 lProAlaProSerSerGluIleProSerIle.ProAla...ProSerGly 1081
1141 CGCAGCCCGAGGCGCTCTCTGCTACTGCGCCACTGGGAGCCAAACCTGC 1190
    |||
    |||
1082 AlAProProValProAlaProSerGlyIleProProValProlYsProSe 1098
1191 A...GCTACACCTGCTGTGCTACAGCCGCGCAACGTGTGAAGCTGCGCC 1237
    |||
    |||
1098 rValAlaAlaIaProProValProlYsProSerValAlaIaValProProValP 1115
1238 CCTGCTGCTTGGGCTGGCGGCCCAACACACTCAAGTCACTGCC... 1284
    |||
    |||
1115 roAlaProSerGlyAlaProProValProlYsProSerValAlaIaAlaPro 1131
1285 .....GCCAGGCGTGGCT.....CTTG 1304
1132 ProValProValProSerGlyAlaProProValProlYsProSerValAl 1148
1305 TGCTGGGAAGTGTAGCTGTGCTGAGCGCTGCCCTGCCACTGCTTCTTACT 1354
    |||
    |||
1148 aAlaProProValProAlaProSerGlyAlaProProValProlYsProS 1165
1355 CCCTC.....CCTGAGCCCTCTTGCCCCCACAA 1383
    |||
    |||
1165 erValAlaAlaAlaProProValProAlaProSerSerGlyIleProProVal 1181
1384 AAGTGCCTGCTGTGCTCTCCCTCTCTCCACCCACCTCAACACTCACATCC 1433
    |||
    |||
1182 ProlYsProAla...AlaGlyValProProValProProProSerGluAla 1197
1434 CCT.....CCATCCTGTAGCTCCCTCAACACAGTGAAGG 1471
    |||
    |||
1198 ProProValProlYsProSerValGlyValProProValProProProSe 1214
1472 TAGAGAGCCACAGTCCCAATCCT 1496
    |||
    |||
1214 rThrAlaProProValProThrPro 1222
```


OW of: US-09-786-136-4 to: A_Geneseq_032802.* out_format : pfs
 Date: Sep 19, 2002 5:41 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=framed_n2p.model -DEV=x1h
 -Q=/cgn2_1/USPTO.spool/US09786136/runat_16092002_141242_19172/app_query.fasta.1.1592
 -DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
 -INS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcpt
 -MAX=100 -THR_MIN=0 -ALIGN=10 -NODE=LOCAL -OUTFMT=pfs
 -M-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09786136.ccg1.1.0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-786-136-4
 Query length: 1611
 Database: A_Geneseq_032802.*
 Database sequences: 747574
 Database length: 111073796
 Search time (sec): 85.420000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA82530	1122.00	1468.77	7.0e-74	21	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW31852	233.00	230.67	6.3e-09	76	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW31855	227.50	286.09	1.5e-08	57	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABB11413	227.00	287.80	1.6e-08	44	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAW72204	222.00	283.45	5.0e-08	31	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03333	220.50	272.79	5.2e-08	90	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW50558	220.00	279.09	5.0e-08	41	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03333	218.50	270.17	7.3e-08	90	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW87503	217.00	265.50	9.9e-08	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW87504	216.50	266.06	1.1e-07	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABW70063	215.00	271.98	1.2e-07	44	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABW71434	215.00	257.12	1.5e-07	22	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG21932	212.00	267.57	2.0e-07	47	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW52322	210.50	263.79	2.6e-07	57	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW67370	210.50	263.79	2.6e-07	57	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW26240	210.50	258.03	2.9e-07	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW26239	210.50	256.84	3.0e-07	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG08112	209.50	257.31	3.4e-07	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABW11527	208.00	255.39	4.4e-07	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW87504	207.50	254.27	4.8e-07	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW87503	207.50	253.06	4.9e-07	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABW11013	207.00	258.71	4.7e-07	60	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AAW55248	205.50	249.64	7.0e-07	13	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABW11013	205.00	256.09	6.6e-07	60	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW28843	205.00	255.58	6.7e-07	64	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW95856	205.00	255.58	6.7e-07	64	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03330	204.50	255.54	7.2e-07	59	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW26240	204.00	249.52	8.6e-07	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW26239	204.00	248.33	8.8e-07	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG14000	201.00	250.97	1.3e-06	59	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03330	200.00	249.65	1.5e-06	59	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG14738	200.00	247.23	1.6e-06	78	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03368	198.50	245.52	2.1e-06	76	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03369	197.00	242.87	2.7e-06	81	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW24516	196.50	225.44	4.0e-06	51	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW38942	195.00	240.90	3.7e-06	76	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW95541	194.50	236.69	4.3e-06	11	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW41231	194.50	227.88	5.0e-06	29	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG23029	194.50	226.03	5.2e-06	38	

alignment_scores:

/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW50363 + 193.50 226.56 6.0e-06
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW50362 + 193.50 226.13 6.0e-06
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG14126 - 193.50 238.62 5.5e-06
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG08321 - 191.50 239.66 6.3e-06
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAW81589 + 191.00 241.83 6.5e-06

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW82530

seq_documentation_block:

ID AAY82530 standard; Protein; 210 AA.

AC AAY82530;

DT 13-JUL-2000 (first entry)

DE Human neurotransmitter associated protein sequence SEQ ID NO:6.

KW Human; neurotransmitter; diagnosis; antiparkinsonian; neuroleptic;
 KW anticonvulsant; nootropic; tranquilizer; neuroprotective; cytostatic;
 KW antidepressant; antidiabetic; gynaecological; immunosuppressive;
 KW neurological disorder; hormone-related disease; Parkinson's disease;
 KW schizophrenia; epilepsy; female reproductive disorder;
 KW attention deficit disorder.

OS Homo sapiens.

XX WO200012685-A2.

XX 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19615.

XX 01-SEP-1998; 98US-0144952.

XX 01-SEP-1998; 98US-0155194.

XX (INCY-) INCYTE PHARM INC.

XX Walker MG, Volkmut W, Klingler TM;

XX WPI; 2000-375619/32.

XX N-PSDB; AAA08405.

PT A new purified polynucleotide comprising a gene that is coexpressed
 with neurotransmitter-processing-specific genes in biological samples
 for diagnosing, treating or preventing neurological and hormone-related
 diseases -

PS Claim 4; Page 35-36; 36pp; English.

CC The present invention describes a purified polynucleotide comprising a
 gene that is coexpressed with one or more neurotransmitter-processing-
 specific genes in biological samples. The neurotransmitter-processing-
 specific genes are L-tyrosine hydroxylase (TH), aromatic amino acid
 decarboxylase (AADC), dopamine beta-hydroxylase (DBH), nicotinic
 acetylcholine receptor alpha3 subunit precursor (nAChR-alpha3),
 secretogranin I and II, Rab3a, human cocaine and amphetamine regulated
 transcript (hcART), vesicular monoamine transporter 1 (hVMAT1), and ARX
 homeodomain protein. The present sequence represents a neurotransmitter
 associated protein sequence from the present invention. The
 polynucleotides from the present invention can have antiparkinsonian,
 neuroleptic, anticonvulsant, nootropic, tranquilizer, neuroprotective,
 cytostatic, antidepressant, antidiabetic, gynaecological and
 immunosuppressive activities. They can be used for diagnosing, treating,
 preventing or evaluating therapies for neurological and hormone-related
 diseases, particularly Parkinson's disease, schizophrenia, epilepsy,
 female reproductive disorders and attention deficit disorder. The gene
 products are therapeutic proteins and targets of therapeutics against
 the diseases.

Sequence 210 AA;

Quality: 1122.00 Length: 210
 Ratio: 5.343 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-786-136-4 x AAY82530 ..

Align seg 1/1 to: AAY82530 from: 1 to: 210

596 ATGGCCCGCCGTCACAACTCTGTAGCCCTGACCGCGCCAGGCCAACAA 645
 |||||
 1 MetGlyArgHisLysLysCysSerProAspProGlyGlnAlaAsnAs 17

646 CAGTGAAGCAGCAGCAGCAGTACTACCTCCCTGGTGGCGGTGCAGAGG 695
 |||||
 17 nSerGlyGlySerAspHisAspTyrLeuProLeuValArgLeuGlnGluA 34

696 CACGAGGCTCTTCGCGCTGACGCGCGCTCTGCGCGCGCTGTGCGCATC 745
 |||||
 34 laProGlySerPheArgLeuAspAlaProPheCysAlaAlaValArgIle 50

746 TCGCAGGAGCGCCCTGTCGCGCTGCGCCCTTGTGCGTGCACCGCGCCAG 795
 |||||
 51 SerGlnGluArgLeuCysArgAlaSerProPheAlaValHisArgAlaSe 67

796 CCTAGCCCGCCACTCGCGCTCATTCGCTGGGCACTTCTGGGCGCTGTG 845
 |||||
 67 rLeuSerProThrSerAlaSerLeuProTrpAlaLeuLeuGlyProGlyV 84

846 TTGGCCAGGGTGACAGTGCCAGCGCTCTCTGCGCGCGCTGCCAGCTCG 895
 |||||
 84 alGlyGlnGlyAspSerAlaThrAlaSerCysSerProSerProSerSer 100

896 GGCTCTGAGGGCCAGCGCAGGTGGACAGTGGCGGGGCTCAGACACCGA 945
 |||||
 101 GlySerGluGlyProGlyGlnValAspSerGlyArgGlySerAspThrGl 117

946 GGCCTCCGATGGGCGGAGGGCTGGGCGGACCGACCTGGGCGGCGCGA 995
 |||||
 117 uAlaSerAspGlyAlaGluGlyLeuGlyGlyThrAspLeuArgGlyArgT 134

996 CTGGGCGCACTGCGCTAGCAGCTGCGCTGGTGGACCGCAGTGGCGCGCT 1045
 |||||
 134 hrTrpAlaThrAlaValAlaLeuAlaTrpLeuGluHisArgCysAlaAla 150

1046 GCCTTCGACGAGTGGGAACCTGACAGCGCGCCAGGCTGATCTGGCTCG 1095
 |||||
 151 AlaPheAspGluTrpGluLeuThrAlaAlaLysAlaAspCysTrpLeuAr 167

1096 GGCCGAGCACTTGCTGACGCGCTGACCTGGCGCGCTCAAGCGCGCAG 1145
 |||||
 167 gAlaGlnHisLeuProAspGlyLeuAspLeuAlaAlaLeuLysAlaAlaA 184

46 CCCGAGGCGCTTCTCTGCTACTGCGCCACTGGGACCAAACTCGCAGCTA 1195
 |||||
 184 laArgGlyLeuPheLeuLeuLeuArgHisTrpAspGlnAsnLeuGlnLeu 200

1196 CACCTGCTGCTACAGCCCGCAGCAACGCTG 1225
 |||||
 201 HisLeuLeuCysTyrSerProAlaAsnVal 210

seq_name: /SIDSL/gcdata/geneseq/geneseq-emb1/AA1997.DAT:AAW31852

seq_documentation_block:
 ID AAW31852 standard; Protein; 763 AA.
 AC AAW31852;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis 74 kDa protein.
 XX
 KW Tuberculosis; mycobacteria; infection; diagnosis;

KW antimycobacterial; antibiotic; vaccine.
 XX Mycobacterium tuberculosis.
 XX WO9741252-A2.
 XX 06-NOV-1997.
 XX 18-APR-1997; 97WO-EP01973.
 XX 29-APR-1996; 96DE-4017184.
 XX (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
 XX Espitia C, Honisch C, Moreno C, Singh M;
 XX WPI: 1997-549750/50.
 XX N-PSDB; AAT93610.
 XX New DNA and related proteins or RNA derived from M. tuberculosis -
 PT used for diagnosis of mycobacterial infections, monitoring
 PT vaccination and development of anti-mycobacterial agents
 XX
 PS Claim 5; Fig 13; 55pp; English.
 XX
 CC This novel 74 kDa protein is encoded by an open reading frame of
 CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
 CC polymorphic GC-rich sequences. Its amino acid sequence shows
 CC a high proline content, but there is no homology to any known
 CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
 CC proteins (see AAW31851-57) are claimed. These can be produced as
 CC recombinant proteins, especially in bacterial, yeast, fungal or
 CC higher eukaryote host cells, and used for diagnosing tuberculosis
 CC and other mycobacterial infections in humans or animals. The
 CC claimed proteins can also be used for epidemiological studies, for
 CC monitoring vaccination, and for the development of vaccines and
 CC anti-mycobacterial drugs.
 XX
 SQ Sequence 763 AA;

alignment_scores:
 Quality: 233.00 Length: 535
 Ratio: 1.131 Gaps: 27
 Percent Similarity: 38.505 Percent Identity: 25.421

alignment_block:
 US-09-786-136-4 x AAW31852 ..

Align seg 1/1 to: AAW31852 from: 1 to: 763

214 CCACAGACTGCTCTTGGTTC.....CACAGGCCATGCCGAGGTGCT 257
 |||||
 23 ProLysSerLysProPheProAlaProAlaProAlaProCysT 39

258 GCGTTCGAGCCCTTCAGACAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 307
 |||||
 39 rpMet.....LeuValSerAlaAlaProCysPro 49

308 ACTTGC.....CC 315
 |||

50 ProAlaProAlaProLysProLysSerLysAlaProPhePro 66
 |||||

316 TGTAGCTGTGGATGCTACTACTAGGAGGTCTCTGCT..... 352
 |||||

66 oValProAlaProAlaArgGluLeuAlaProProLeuProProA 83
 |||||

353GGGCGCTGCAGGTGTC... 370
 |||||

83 laProProGluAlaProArgGluSerArgProAlaLeuProCysPro 99
 |||||

371AGCTCAGAGCCGCTGAGCCCGCCAGG 396
 |||||

```

100 ProProValValIleProAspProProGluProAlaAlaProPro... 115
397 ACCCCCTCTGCTCTCAGC... 418
116 .ValProProAlaProAsnSerProProPheProProAlaAlaP 132
419 ....CATCTAGATGAGTCTCTGCGGACCTGTTGCTACTCTAAAGGA 463
132 roLysPheValProAlaProProValProProValProAsnSerProPro 148
464 CTTACAGAGAGGCTCTCAGCAGCGGCTGGGACTCGGACCAAAATGGCAA 513
149 PheProProPheProAla... 155
514 CTCACAGCTGCTTTGGGGAGCCCTGCCACTCCACGGAAGTCTCGCC 563
156 .....AlaLeuAsnProProAlaProProAlaProProLeuAla 169
564 GCCACCTCCCGCTCTCCCTCT... 586
169 snSerProProLeuProProAlaProProThrProAlaGlyThrProPro 185
587 .....CGGCTCAGCATGGCGCGCTCACAACACTCTGTAG 621
186 AlaAlaProTrpProProValProAlaAlaProLysSerLysProAlaSe 202
622 CCTGTACCGCGGAGGCAACAGCATGAGGACGACGACCATGACTACC 671
202 rProProArgProProAlaProProMetProAlaThrProMetGluPheP 219
672 TGCCCTTG.....GTGCGGCTGCAGGAGGACCA 700
219 roProLeuProProValProProAspProProLysGluThrProPro 235
701 GGCTCTTCGCTGAGCGCGCTTCTGCGCGCTGTGCGCATCTCGCA 750
236 .AlaProProAlaProProIleProProAlaProVal..... 247
751 GGAGCCCTCTCGGCTGCTGCTGCGGCTGCTGCGGCTGCTGCTGCTG 800
248 ..ProIleProProAlaProPro...ProLeuPro...ProValProAsnLys 261
801 GCCCACCCTCGGCTCATTTGCTGCGGCTGCTGCGGCTGCTGCTGCTG 850
262 IleProProAlaProProAlaProProValAlaAlaAlaValLeuVa 278
851 CAGGGTGACAGTGCCACGCGCTCTGACGCGGT.....C 885
278 lAlaProCysProProProLeuProProAlaProAsnHisProProAlaP 295
886 CCCCAGCTCGGCTCTGAGGGGCGGAGGCGGAGTGGACAGTGGCGGGGCT 935
295 roProAlaAlaProValProGlyValProLeuAlaProLeuProAsnSer 311
936 CAGACACCGAGGCTCCGATGGGCGGAGGAGGCTGGGCGGACCGACCTG 985
312 HisProProAlaProPro.....SerAlaProValP 322
986 CGGGCGGACCTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1016
322 oGlyValProLeuAlaProLeuProIleSerGlyArgProValSerValT 339
1016 ..... 1016
339 rPLysGlySerPheThrThrLeuSerThrPheCysCysArgValCysSer 355
1017 .....TCGCTGCTGGAGCAGCAGTGGCGC 1042
356 GlyGluValLeuAlaGlyAlaLeuAsnProSerArgProSerArgSerPr 372
1043 GCTGCTTCGACGAGTGGGAAGTACAGCGGCGGCGGCTGATGCTGGCT 1092
372 oLeuThrThrThr.....ProAlaLeuProAlaProI 384

```

```

1093 CGGGCCCGCAGCAGCTGCTGCTGACGGCCTTG.....ACCTGGCCG 1130
384 leProProLeuProProLeuProProLeuProIleAsnThrAlaValPro 400
1131 CCCTCAGGGCGCAGCCCGAGGGCT.....CTTCTTG 1162
401 ProIleProProLeuProPro.ValThrAlaLeuAlaProProLeuPro 417
1163 CTACTGGCCCACTGGACCAAAACCTGCAGCTACACCTGCT...GTGCTA 1209
417 roLeuAlaProLeuProIleSerProGlyValProProAlaProProIle 433
1210 CAGCCGAGCGAAGCTGGAAGGTCGCCCTGCTGCTGCTGCTGCTGCTG 1259
434 ProProGlyLysProTrpThrThr.ProProLeuAlaProAlaProPro 450
1260 CACCACACACTCAAGTCACTGCGCGGCGGCTGCTGCTGCTGCTGCTG 1309
450 luProLysThrValProValLeuProProGly..... 460
1310 GGAAGTGTAGGCTGGTCCAGCCTGTCCCGCCTG...CTTCTTACTCC 1356
461 .....ProSerCysProProSerGluLysProAsnPr 471
1357 CTCCCTGGAGCCTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1406
471 oProAlaProProGluProProGluProLysSerSerProAlaLeuPro 488
1407 CCTCTCTCCCGCAGCCTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1453
488 roAlaProProAlaPro...SerMetProSerAlaValArgValPro 502

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW31855

seq_documentation_block:

ID AAW31855 standard; Protein; 572 AA.

XX AC AAW31855;

XX DT 27-APR-1998 (first entry)

XX DE Mycobacterium tuberculosis 55 kDa protein.

XX XX Tuberculosis; mycobacteria; infection; diagnosis;

KW antimycobacterial; antibiotic; vaccine.

XX OS Mycobacterium tuberculosis.

XX PN W09741252-A2.

XX PD 06-NOV-1997.

XX PF 18-APR-1997; 97WO-EP01973.

XX PR 29-APR-1996; 96DE-4017184.

XX (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.

PA Espitia C, Honisch C, Moreno C, Singh M;

XX WPI; 1997-549750/50.

DR N-PSDB; AAT93610.

XX New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents

XX PS Claim 11; Fig 16; 55pp; English.

XX This novel 55 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows


```

171 roSerProAlaAlaSerProAlaPro...ProAlaAlaSerProValLeu 186
1029 AGCACCGATGGCGGCTTGCCTTCGACAGCTGGGAAGTGCACAGGCGCAAG 1078
187 ThrAlaSerProLeuProAlaAlaSer.....Prose 198
1079 GCTGATTGCTGGCGGGCCAGCAGCTTCCTGAGCGGCTTGACCTGGC 1128
198 rProAlaAlaSerProAlaProProAlaAlaSerProValLeuThrAla 215
1129 CGCCCTCAAGGCGCAGCCGAGGGCTCTTCCTGCTACTGCGCAGCTGG 1178
215 erProLeuProAlaAlaAlaSerProSer..ProAlaAlaSerProAlaPr 231
1179 ACCAAAA.....CTGCGACGTA 1195
231 oProAlaAlaSerProValLeuThrAlaSerProProLeuProAlaAla 248
1196 CACCTGCTGTGTACAGCCCGAGGAGCTGTGAAGCTGCCCGCTGCTGC 1245
248 erProAlaAlaAlaAlaSerProValHisThrAlaSer..ProProValH 264
1246 TTGGGCTGGCGCCCAACACACACTCAAGTCACATGCGCGCCAGGCGTG 1295
264 sValAlaSerProProValHisThr.....AlaSerProPro..... 276
1296 GCCTCTTTGGTGTGGGAAAGTGTAGGCTGGTGCACGCTGTCCCCACATG 1345
277 ....ValHisValAlaAlaSerProProValHisThrAlaSerProProVal 291
1346 ...CTTCTTACTCCCTCCCTGGAGCCCTCTTGCGCCCAACAAAAGTGCCT 1392
292 HisValAlaAlaSerProProValHisThrAlaAlaSerProHisValHisValAl 308
1393 G.....CCTGTGCTCTCTCCCTCTCCTCCCGCCCACTCACACTCCCT 1436
308 aSerProProProValHisValAlaAlaSerProProVal..HisValAlaSerPro 324
1437 CCA 1439
325 Pro 325

```


CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (ii) and its binding partners are useful in medical
 CC imaging of sites expressing (iii). (i) and (ii) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 900 AA;

alignment_scores:

Quality: 220.50 Length: 551
 Ratio: 0.930 Gaps: 35
 Percent Similarity: 43.013 Percent Identity: 27.042

alignment_block:

us-09-786-136-4 x ABG03533

Align seg 1/1 to: ABG03533 from: 1 to: 900

```

201 GGTGGTGGGGCTCCACAGACTGCTCTGGTTCCACAGCCATGCCCG 250
||| ||||| ||||| ||||| |||||
177 GlyProGlySerProArgCysAsnArgCysArgGluArgLysProG 193
251 A...GGTGGTGGCTCCAGGCC...TTCACACAAGTAGTCACTGTC 294
||| ||||| ||||| ||||| |||||
193 yThrProGlyTrpProArgLeuArgSerProGlyAsnLeuArgP 210
295 CCCCTCTGCTTCACTGCTGCTAGCTGTGGATGCTACTACTAGGAG 344
||| ||||| ||||| ||||| |||||
210 roGlyValGlyLeuGlyLeuAlaLeu..... 219
345 TCCTGGCTGGGGCTCCAGCTGTGACGTGACAGCCCGCTGAGCCC... 391
||| ||||| ||||| ||||| |||||
220 .....ProAlaArgThrAlaAlaAlaPro...ArgProArg 231
392 .....CCAGAACCCCTCCTGCTCTCACAGCCATCTAGA 426
||| ||||| ||||| ||||| |||||
232 GluArgTrpArgSerProGlyAlaProCysLeuGly.....AlaG 245
427 TGCAGCTCCTGCTGCCACTGTGTCTACTCTAAGGACTTCAGAGAGG 476
||| ||||| ||||| ||||| |||||
245 n**ProSerLeuProAlaAlaGluProArgGly.....ProGlyT 260
477 CTCAGCAGCGCTGGGACTCGGACCAAAATGGCAACTCCAAG...CGT 523
||| ||||| ||||| ||||| |||||
360 hrSerGlyArgAlaTrpProSerSerAlaAlaArgGluaspCysArg 276
524 GCCTTGGGG...GACCTGCACTCCAGGAGAGTCTCGCGCCGCCACC 570
||| ||||| ||||| ||||| |||||
277 AlaProGlyArgGlyProAlaAlaProThr...GlyAlaArgArg...Pr 291
571 TCCCGCTCCTCCCTGCTGGGCTCAGCATGGGCCCGCTCACAAACTCTG 620
||| ||||| ||||| ||||| |||||
291 OProArgProGlyAlaGlyLeuAla..... 299
621 GCCCTGACCCGGCCGACCAACACAGTGAAGGCGCAGCCACTGACTAC 670
||| ||||| ||||| ||||| |||||
300 .....GlyArgProThrArgAlaArgAla..... 307
671 CTGCGCTTGGTGGCTGACAGGAGGACCAAGGCTCTCTCGCGCTGACGC 720
||| ||||| ||||| ||||| |||||
308 LeuProGlyArgIleGlyArgArgSerProGlyArgAlaAlaProHis 324
721 GCCCTTCTGCGCGCTGCTGGCGCATCTCGCAGGAGCGCTCTGCGCGCT 770
||| ||||| ||||| ||||| |||||

```

```

324 exProAlaGlyLeuArgSerAlaAlaArgArg.....CysPro 336
771 CGCCCTTTCCCTGACCCGCGCAGCCTCAGCCCGCCTCGGCCTCATTTG 820
||| ||||| ||||| ||||| |||||
337 ProProGlyProCys...AlaProArgGlyGlnProGlyLeuProArgG 352
821 CCTTGGG.....CACTTCTGGGCTCTGGTGTGGCCAGG 855
||| ||||| ||||| ||||| |||||
352 yGlyGlyAlaAlaProLeuArgHisGlyGlyAlaGlnLysThrArgP 369
856 TGACAGTGCACAGGCTCTCGCAGCCCGTCCCGCAGCTCGGCT..... 899
||| ||||| ||||| ||||| |||||
369 roSerGlnGlyArgProGlyProProProProAlaProAlaGlyGlu 385
900 .....CT 901
386 ArgGlnProGlnProGlyLysGluSerGlyGluLysGlyValGlnProAl 402
902 GAGGGGCCAGGCGAGGTGACAGTGGGGGGCTCAGACACGAGGCCTC 951
||| ||||| ||||| ||||| |||||
402 aArgGlyGluProGlnGlyProGlnArgAlaMetArg..... 414
952 CGATGGGCGGAAGGCTGGGCGGCACCGCCTCGGGGGCGGCGGCTGGG 1001
||| ||||| ||||| ||||| |||||
415 .....GlyAlaProAlaGlyArgArgProProAlaProAlaProG 429
1002 CCACCTGCCCTAGCCTCGCTGGCTGGAGCACCAGTGGCGGCTGCCTTC 1051
||| ||||| ||||| ||||| |||||
429 y.....AlaAlaAlaGlnPro.....G 435
1052 GACGAGTGGGAAGTACAGCGCCGAGGCTGATGCTGGCTGGCGGCCCA 1101
||| ||||| ||||| ||||| |||||
435 lYargTrpGlyThrAlaAlaGly.....AlaAlaGlyPro 446
1102 GCACCTGCTGACGGCTGACCTGCGCCGCTCAAGCGCGCAGCCCGAG 1151
||| ||||| ||||| ||||| |||||
447 Ala...AlaGlyArgProAlaProGlyArgGlnProSerArgAlaAla 462
1152 GCCTCTCTCTGCTACTGGCCCA..... 1173
||| ||||| ||||| ||||| |||||
462 uGly...ProAlaAlaAlaProGlyGluArgGlyGlnSerIleProValP 478
1174 .....CTGGGACCAANA 1185
||| ||||| ||||| |||||
478 roIleGlyValValGluGluArgGluGlyProArgGlyProGlyProThr 494
1186 CTGAGCTACAGCTGCTGCTGACAGCCCGCAGCACTGTCAAGCTGC 1235
||| ||||| ||||| ||||| |||||
495 AlaAlaAlaGlyProAlaAlaProGlyArgGlyArgAlaGlySerAl 511
1236 C.....CCCTGCTGCTGGGCTGGCGCCCGCAGCCCAAC 1267
||| ||||| ||||| ||||| |||||
511 aArgAlaArgGlyProGluProAlaAlaProAlaArgGlnProGlyArg 528
1268 ACACCTCAGTCACTGCGCCCGCAGGCG..... 1293
||| ||||| ||||| ||||| |||||
528 rgAlaArgLeuProArgThrArgGlyProProArgProAlaProSerGlu 544
1294 ...TGCCCTCTT.....GGTCTGGGAAAGTGTAGGCTGG 1325
||| ||||| ||||| ||||| |||||
545 ProTrpProLeuProSerArgProProGlySerGlyLys.....ArgG 559
1326 TCCAGCCTGTCTCCCC.....ACTGCTCTTACTCTCC 1357
||| ||||| ||||| ||||| |||||
559 yAlaSerAlaProProGlyGlyCysProSerArgSerGlyAsnSerPro 575
1358 TCCTTGGAGCCCTCTTGGCCCGCAGCAAAAGTCCCTGCTGCTCTCTCC 1407
||| ||||| ||||| ||||| |||||
576 ThrLeuGlnAspProGlyProHisSerGlyCys...ProGlu.GlyPheP 591
1408 CTCTCTCTCCCGCAGCTCAGCTCC..... 1433
||| ||||| ||||| ||||| |||||
591 roArgLeuAlaGlyGlyGluArgSerSerGlyProGluMetGlyThrVal 607

```


259ProProProLeuProValSerSerProSerPheProProPro 273
811 GGCCTCATTTGCCCT.....GGGCACTTC 833
DR |||||
273 hrSerProCysProArgGlyAlaSerAlaGlyProValGlyGly 289
PT TGGGCCCTGGTTGGCCAGGTGACAGTGCCA..... 866
PT :||:|||||:|:|:|
290 PheLeuArgAlaAlaSerArgValLeuAlaProProThrProArgGlyPr 306
XX
867CGGGCTCTCGAGCCCGCTGCCAGCTCGGGCTCGA 903
PS ||||| |||||:|:|:|
306 oArgGlyArgArgProAlaArgAlaArgProAlaProGluThrThr 323
CC ||||| |||||
904 GGGCCAGCCAGGTGACAGTGGCGGGGCTCAGACACGAGCCCTCCG 953
CC ||:|||||:|:|:|
323 lyGly.....AlaThrGlyGlyArgGlySerProAlaArgPro 335
CC ||||| |||||
954 ATGGGGGAAGGCTGGGGGACCCAGCTGGGGCCGAGCTGGGCC 1003
CC ||||| |||||
336 ProGlyCysProGlyGlnArgAla...CysCysProProGly..... 348
CC ||||| |||||
004 ACTGCCCTAGCAGCTGGCTGGGACCCAGTGGCGCGCTGCTTCGA 1053
CC ||||| |||||
349SerSerArgGlyCysSerThr.....ProThrArgGlyA 360
CC ||||| |||||
1054 CGAGTGGGAAGTACACAGCGGCCAGGCTGATT.....GCT 1088
CC ||||| |||||
360 laSerGlySerArgGlyAlaSerArgSerCysAlaArgAlaArgGlyAla 376
CC ||||| |||||
1089 GCTTGGCGGGCCAGCAGCTTGCCTGACGCGCTTGACCTGGCGGCCCTCAAG 1138
CC ||||| |||||
377 GlyCysGlyGlySerThr.....ArgG1 384
CC ||||| |||||
1139 GCCGAGCCCGAGGGCTCTCTGCTACTGCGCCACTGGGACCAAAACCT 1188
CC ||||| |||||
384 yProArg, ProAlaAlaGlyProProArgSerProSerGlyProArgGly 400
CC ||||| |||||
1189 GCAGCTACAGCTGTGTGCTACGCCAGGACGAGTGTGAAGGCTGCCCC 1238
CC ||||| |||||
401 ProArgArgProLeuCys...GlyProSer.....Proc 411
CC ||||| |||||
1239 CTGCTGCTGGCGTGGCGCCGCCACCC 1264
XX :||:|||||:|:|:|
411 ysProGlyArgAlaGlySerProPro 419

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG03533

seq_documentation_block:

ID ABG03533 standard; Protein; 900 AA.

XX

ABG03533;

3-FEB-2002 (first entry)

DE Novel human diagnostic protein #3524.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSB-) HYSSEQ INC.

XX

XX

XX

XX

XX

XX

PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS67720.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
Claim 20; SEQ ID No 33892; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 900 AA;

alignment_scores:

Quality: 218.50 Length: 593

Ratio: 0.892 Gaps: 38

Percent Similarity: 41.315 Percent Identity: 26.813

alignment_block:

US-09-786-136-4/rev x ABG03533 ..

Align seg 1/1 to: ABG03533 from: 1 to: 900

1411 AGAGGGAGAGACAGCAGCGACCTTTTGTGGGCAAGAGGCTCCA 1362

||| :||:|||||

29 ArgProAlaAlaGlnLysGlnAlaLeu..... 37

1361 GGGAGGGAGTAAGAAGCAGCTGGGGACAGGCTGGCACCAGCCTACACTTT 1312

|||||:|:|:|:|

38GlySerArgGluArgValGlyThrGly..... 46

1311 CCCAGCACCAAGAGCCACCTCGGGGCACTGACTTGAGTGTGTGGG 1262

|||||:|:|:|:|

47ProGlyArgIle...LeuArgProGlyGly 55

1261 TGGGGGCCAGCCCAAGCAGCAGGGGAGCCCTTCACACGTTCCGCTGGGC 1212

|||||:|:|:|:|

56 TrpGlyCysPheProGlyProArgGlyThrGluAspAlaAspGlnArgAl 72

1211 TGTAGCACAGCAGGTGTAGCTGCAGGTGTTT..... 1182

|:|:|:|:|

72 a..AlaArgGlyProValGlyAlaGlyThrGlnGlnHisGlyArgAlaVal 88

1181GGTC 1178

88 lProArgGlyProGlnAsnGluProAspGluThrLeuLeuProGlyGlyP 105

1177 CCAGT.....GGCCAGTAGCAGGAAGACC..... 1152

||||| ||| :|:|:|:|

```

105 roSerProArgGlyGlyGluLeuArgGlyArgSerGlyAlaArgGlyLeu 121
1151 .....CTCGGCTCGGCTTGGAGCGCGCCAGGTCAAGCGC 1114
122 Pro***SerLeuThrGlyProAlaProGlyProGlnArgGlyGly** 138
1113 TCAGCAAGTGTGGCCCGCAGCAGCAATCAG...CCTGGCGCTGT 1067
138 *SerProSerProGlyArgAlaSerSerLysAlaGlyProTrpLysArgP 155
1066 CAGTTCCTCCACTCGTCCGAAGG...CAGCGCGCATCGGTCCACCCAGG 1020
155 roGlyAlaSerArgAlaSerLeuGlnArgAlaSerSerMetProAlaSer 171
1019 CGAGTGCTACGCAAGTGGCCCGAGTCCGCCCGCCGAGT..... 981
172 GlnVal.....AspTrp.....GlyGlyProGlyGlySerProArgCy 184
980 ....CGGTCCGC.....CCAGCCCTTCCGCCCGCCATCGGAGCCCTC 944
184 sasnArgCysArgGluArgLysProGlyThrGlyProGlyTrpProProA 201
943 GGTGTGTAGCCCGCCCACTGTCACCTGGCTGGCCCGCCCTCAGAGCCCG 894
201 rgLeuArgSerProGlyAsnLeuArgProGlyValGly..... 213
893 AGCTGGGGACGGGTGTCAGGAGCGCG..... 867
214 .....GlyLeuGlyLeuAlaLeuProAlaArgThrAlaAlaAlaApr 228
866 .....TGCCACTGTCCACCTGGCCAA.....C 845
228 oArgProArgGluArgTrpArg...SerProGlyAlaProCysLeuGlyA 244
844 ACCAGGCCCAAGTGGCCCGGCAAGTGGCCCGAGGTGG..... 804
244 laGln**ProSerLeuProAlaAlaAlaGluProArgGlyProGlyThr 260
803 ...GGCTGAGGTGG..... 792
261 SerGlyArgAlaTrpProSerSerAlaAlaAlaArgGluAspCysArgAl 277
791 .....CGCGTGCAGCGCAAGGGCGAGGC 767
277 aProGlyArgGlyProAlaAlaProThrGlyAlaArgArgProProArgP 294
766 ACCGAGAGCGGCTCTCGCAGATGCGCAGCGCGCAGAGGGCGGT 717
294 roGlyAlaGlyLeuAlaGlyArg.....ProThrArgAlaArgAlaLeu 308
716 CCAGGC.....GGAGAGCGCTGGTCTCTGCGCGCGCACCAGG... 675
309 ProGlyArgIleGlyArgArg...SerProGlyArgAlaAlaProHisSe 324
674 ....GCAGTAGTCATGGTGGCTGCTTACATGTTGTTGGCTGGCCGG 629
324 rProAlaGlyLeuArgSerAlaAlaArgCysProProGlyPro. 340
628 GTCAGGCTACAGAGTTTGTGA.....CGCGCGCCATGCTGAGCCGA 586
341 .....CysAlaProArgGlyGlnProGlyLeuProArg 351
585 CAGGGAGGA.....CGGGAGGTGG.....CGCGAGG 557
352 GlyGlyGlyAlaAlaProLeuArgHisGlyGlyAlaGlnLysThrAr 368
556 ACCTTCGTGGAGTGGCGAGGTGCCCCCAAGCAGCTTGGAGTTGCCAT 507
368 gProSerGlnGlyArgProGlyPro..... 377
506 TTTGGTCCGAGTCCAGGCGCTGCTGGAGAGCTCTCTGAAGTCTTTA 457
378 .....ProProProAlaAlaProAlaGlyGluArgGlnProGlnPro... 390

```

```

456 GAGTAGACAAGTGGCGCAGAGGAGCTGCATCTAGATGGCTGTGAGAGC 407
391 .....GlyLysGluSerGlyGluLysGlyValGlnProAl 402
406 AGGAGGGTTCTCT...GGGGCTCAGCGGCTCTGAGCTGCACACCTGCA 360
402 aArgGlyGluProGlnGly.ProGlnArgAlaMetArgGlyAlaProAla 418
359 GGG.....CCCCAGGAGGACCTCCCTAGTAGTAGTATCCACAGCT 319
419 GlyArgArgArgProProAla.ProAlaProGlyAlaAlaGlnProG 435
318 ACAGGGCAAGTGAAGCAGGAGGGGAGCTGACCTTACTTGTCTGAAG 269
435 lyArgTrpGlyThrAlaAlaGlyAlaAlaGlyPro.....Ala 447
268 GGCTCGAAGCCAGCACCT.....CC 249
448 AlaGlyArgProAlaProGlyArgGlnProSerArgAlaAlaGluGlyPr 464
248 GGGCATGGCTGTGGAAACCAAGAGCAGTCTGTGGAGCCGCCACCCCA 199
464 oAlaAlaAlaProGlyGluArgGlyGlnSerIleProValPro..... 478
198 CCCTCTCAGAGGCTCATAGAAAA..... 173
479 .....IleGlyValValGluGluArgGluGlyProArgGlyProGly 492
172 ...ATTGAGTGTCTCCCGAGCTTCTCCCTCTATCGACTTTACAGAGT 126
493 ProThrAlaAlaAlaGlyProAlaAlaPro.....GlyArgG 505
125 AGATTGCTTGTAGGCGCCAGACAGCTCAGACTCCCACTTTGACAA 76
505 yArgArgAlaGlySerAlaAlaArgGlyProGluProAlaAlaProA 522
75 GAATGTATCAGGCTGAATCTTCAACCCATAGCAGAGCAAGCCCTGG 26
522 laArgGlnProGlyArgArgAlaArgLeuProArgThrArgGlyProPro 538
25 CTGCCAGAAATTCAGCAGCTCTGG 2
539 ArgProAlaProSerGluProTrp 546

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW87503
seq_documentation_block:
ID AAW87503 standard; Protein; 1212 AA.
XX
AC AAW87503;
XX
DT 23-FEB-1999 (first entry)
XX
DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
XX
KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW NMDA-activated cation-selective ion channel; glutamate receptor.
XX
OS Homo sapiens.
XX
PN US5849895-A.
XX
PD 15-DEC-1998.
XX
PF 20-APR-1994; 94US-02311193.
XX
PR 20-APR-1994; 94US-02311193.
XX
PR 20-APR-1993; 93US-0052449.
XX
PA (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI Daggett LP, Lu C;

```



```

1019 AlaArgCysProLeuGlyAlaProAla...Pro.ProAlaProAlaProT 1034
324 ACAGCTACAGGCAAGTGAAGCAGAGGGGGCAGAGCTGACCTTACTGTGT 275
1034 hrAlaThrArgProAlaGlyAlaTTPArgArgSerArgCysAlaCys 1050
274 CTGAAGGGCTCCGAGCCAGCA.....CTTCC 249
1051 ArgSerThrGlyArgProAlaAlaArgAlaSerArgGlnGlyProProPr 1067
248 GGGCATGGCTCTGGAACCAAGGAGCAGCTGTGTGGAGCCGCCACACCCA 199
1067 oGly.....SerThrAspSerThrSerAlaCysThrProThrProT 1081
198 CCCTCTAC.....AGAGGGCTC.....ATAGAAAAAATGCA 167
081 hrCysHisCysAlaGlyGlyLeuSerValLeuThrPheHisProValThr 1097
166 GTGTTCTCCCGAGCTCTCCCTCT 143
1098 AlaThrAlaProGlySerProAla 1105

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW87504

seq_documentation_block:

ID_AAW87504 standard; Protein: 1061 AA.

XX AC AAW87504;

XX DT 23-FEB-1999 (first entry)

XX DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.

XX KW Human; N-methyl-D-aspartate receptor; NMDAR2C;

XX KW NMDA-activated cation-selective ion channel; glutamate receptor.

XX OS Homo sapiens.

XX PN US5849895-A.

XX PD 15-DEC-1998.

XX PF 20-APR-1994; 94US-0231193.

XX PR 20-APR-1994; 94US-0231193.

XX PR 20-APR-1993; 93US-0052449.

(SIBI-) SIBIA NEUROSCIENCES INC.

PI Daggett LP, Lu C;

XX WPI; 1999-069812/06.

XX DR N-PSDB; AAW82910.

XX PT DNA encoding N-methyl-D-aspartate receptor subunit - useful for the assembly of functional glutamate receptor subunits

XX PS Example 3; Columns 261-268; 203pp; English.

XX The present sequence represents a human N-methyl-D-aspartate (NMDA) receptor subunit (NMDAR). The nucleic acid sequence does not contain the 860 5'-most nucleotides, has an additional 11 nucleotides (AAV82891) between nucleotides 1300 and 1301, an additional 24 nucleotides (AAV82890) inserted between nucleotides 23050 and 2351, as set forth in AAV82889. The cDNA sequence is derived from clone NMDA24. The NMDAR subunits contribute to the formation of NMDA-activated cation-selective ion channels. In addition to being useful for the production of NMDA receptor subunit proteins, the nucleic acids are also useful as probes to identify and isolate nucleic acids encoding related receptor subunits. Functional glutamate receptors can be assembled from several NMDA receptor subunit proteins of one type (homomeric) or from combinations of subunit proteins of different types (heteromeric). The present invention

CC also comprises methods for using such receptor subunits to identify and CC characterise compounds which affect the function of such receptors, e.g. CC agonists, antagonists and modulators of glutamate receptor function. The CC invention also comprises methods for determining whether unknown CC protein(s) are functional as NMDA receptor subunits.

XX Sequence 1061 AA;

alignment_scores:

Quality: 217.00 Length: 528

Ratio: 0.995 Gaps: 36

Percent Similarity: 41.288 Percent Identity: 28.220

alignment_block:

US-09-786-136-4/rev x AAW87504 ..

Align seg 1/1 to: AAW87504 from: 1 to: 1061

```

1487 GGACTGTGGCTCTCTACCTTCCACTG.....TGTTCGAG 1453
||||| ||| ||||||||| ||| ||
526 GlyLeuAlaArgSerLeuProLeuAlaThrAlaSerProCysArgAr 542
1452 GGAGCTCAGAGCA.....TGGAGGGAGTGT...GAGTGGGT 1418
| ||| ||||| ||||| ||||| |||||
542 gThrProThrGlySerGlyProThrTrpArg.SerCysSerSerTrpGly 558
1417 GGGAGGAGGAGGAGCAGCAGCAGCAGCTTTTGTGGGGGCAAGAGG 1368
:||||: |||||: |||||: ||
559 ThrGluArgHisArgAsnTrpArgGlnCys.....G1 569
1367 CTCTCAGGGAGGAGTAAAGAGCAG..... 1343
| |||||||: |||||: |||||: |||||
569 ySerGlnGlySerAlaArgMetArgThrArgAlaAlaSerTrpThr 586
1342 ..... 1328
586 erThrThrTrpGluAlaSerSerThrCysCysTrpTrpProTrpGlyTrp 602
1327 CACCAGCTTACACTTCCAGCACCAAGAGCCCTCGGGGGCAGTG 1278
||| ||| ||||| |||||
603 .....ProCysTrpSer.....SerProGlySerThr... 611
1277 ACTTGAGTGTGTTGGTGGGGGGCCAGCCCAAGCAGCAGGGGGCAGCTT 1228
|||||: |||||: |||||: |||||
612 .....TrpSerThrGlySerCysAlaThrArgCysProT 623
1227 CACAGTTCGCTGGGCTGT.....AGCAGCAGAGGTAGCT 1191
:||||: |||||: |||||: |||||: |||||
623 hrHisProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 639
1190 GCAGGTTTGTCTCCAGTGGCGCAGTAGCAGGAAGAGCCCTCGGGGCTCG 1141
|||||: |||: |||||: |||||
640 AlaSerAlaGlyCysArgAlaSerProAlaHisArg..... 651
1140 GCCTTGAGGGGGCCAGGTCAA...GGCGTCAGGCAAGTCTCGGCCCG 1094
|||||: |||||: |||||: |||||
652 .....GlyArgProAlaArgThrSerArgProAlaArgProArgProA 666
1093 CAGCAGC.....AATCAGCT.....TGGCGGCTGTCAGTTCCC 1059
|||||: |||||: |||||: |||||
666 laCysSerArgPheCysArgGlnProAlaThrTrpPro..... 678
1058 ACTCGTCGAAGGAGCGCGCATCGGTCTCCAGCCAGCGAGGTGCTACG 1009
679 .....ArgArgAlaAlaTh 683
1008 GCAGTGG.....CCAGGTCCGGCCCCCAGGTGGGT 977
||| ||| ||| |||||
683 rProTrpThrAlaProLeuAlaProSerArgIleGlyValAlaAlaLav 700
976 GCGCGCCAGCCCTTCCGCCCATCGAGGCTCGGTGCTCAGGCCGCC 927
||||| ||||| ||||| |||||

```

700 alArgProHisIArgProPro...AlaArgProArgGlyLeuAlaProAla 711
926 CAC...TGTCCACCTGGCGCTGGCCCTCAGAGCCGAGCTGGGGAGCGG 880
||| ||||||||| |||||||||
716 HisAlaCyysProProThrArgProGlnSer..... 726
879 CTGCAGGAGCGGTGGCACTGTCACCCTGCCAACACCAGGGCCCAAGA 830
726 726
829 TGCCCNAGGCCAATGAGCGGAGTGCGGCTGAGGCTGGCGCGGTGCACGG 780
||||| ||||| ||| |||||
727ArgAla.....ProArgAlaGlyAspArgGlnThrGlyValAla 739
779 CAAGGGCGAGCAGCGGAGAGGCGCTCGCGAGATGGCACAGCGGG 730
:::||| ||| :::||| ||| ||| |||
740 ArgArgLeuCysAlaGlyLeuArgSerProArg.....AlaAlaProAr 754
729 CAGAGGGCGGCTTCAGCGGAGGAGCGCTGGTGCTCTCTCGACGCGCAC 680
||||| :::||| ||| |||
754 gArgArgGlyArgPro.....CysProThrSerProGluC 766
679 CAAGGCGAGGTAGTCATGCTGCTCGCTCATCTGTTGTTGGCGCTGCCGG 630
||||| :::||| ||| ||| |||||
766 ysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGlyProGlyThr 782
629 GGTACGGGCTACAGAGTTGTGACGGCGGCCCATGTGAGCGCACAGGGA 580
:::||| ||| :::||| ||||| :::||| |||||
783 AlaGlyGlyThrSerArgProProSerGlyProCys..ArgProArgAlaV 799
579 GGACGGGGAGG.....TGGCGCGCAGGAGCTTC 551
||| ||||| :::|||
799 alThrThrAlaProPheLeuGluProThrAspProAlaAlaProSerSer 815
550 CGTGGGAGTGGCGAGGTCCTCCCCAAAGCAGCCTGGAGTTGCCATTGTTGT 501
||| ||||| ||| ||||| ||| |||||
816 Arg...SerSerArgSerProArgSerTrpArgThrCysArgCysSerVa 831
500 CGAGTCCCAGCGCGCTCTCGAGAGCCTCTCTGAAGTCCTTAGAGTAG 451
||||| ||||| :::||| ||||| ||||| |||||
831 lArg.SerSerTrpProGlyArgProCys..... 841
450 ACNACAGTGGGCGAGGAGCTGCATCTAGA..... 421
||| ||| ||||| |||||
842 ThrArgProGlyProGlyAlaArgAlaArgValThrLeuProcysProAl 858
420TGGCTGTGAGAGGAGGAGGGTTCTGGGGCTCAGCGGGCTCTG 375
||| ||| ||||| ||||| |||||
858 aProTrp.....ProArgProSerLeuGlyPro. 867
374 AGCTGCACACCTGCAGGGCCCCAGCAGGAGCTCTCCCTAGTAGTACATCC 325
:::||| ||| ||| ||| ||| ||||| :::|||
868 AlaArgCysProLeuGlyAlaProAla...Pro.ProAlaProAlaProTr 883
324 ACAGCTACAGGCAAGTGAAGCAGGAGGGGCGCAGCTGACCTTACTTGT 275
:::||| ||| ||| ||||| |||||
883 hrAlaThrArgProAlaGlyAlaTrpargargSerArgCysAlaCyys 899
274 CTGAAGGGCTGCAAGCCACGA.....CCTCC 249
:::||| ||||| ||||| |||||
900 ArgSerThrGlyArgProAlaArgAlaSerArgGlnGlyProProPr 916
248 GGCATGGCCTGTGGAACAAGGAGCAGCTCTGTGGAGCCCCCACCACCCA 199
||||| :::||| ||| ||||| |||||
916 oGly.....SerThrAspSerThrSerAlaCysThrProThrProt 930
198 CCCTCTAC.....AGAGGGGCTC.....ATAGAAAAAATTCGA 167
||| ||||| |||||
930 hrcyshiCysAlaGlyGlyLeuSerValLeuThrPheHisProValThr 946
166 GTGTTCTCCCCAGCTTCCTCCCTCT 143
:::||| ||||| ||||| |||||
947 AlAlaThrAlaProGlySerProAla 954


```

39 hrSerGlnProProProSerSerIleGlyLeuSerProProSerAlaPro 55
392 CCAGGACCCCTCCTGCTCTCAGACCCATCTAGATGACGCTCCTGCTGCC 441
56 ThrThrProPro...SerArgGlyHisVal...ProSerPr 68
442 CACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGCGCT 491
68 o.....ArgHisAlaProProArgHisAlaAt 77
492 GGGACTCGGACCAAAATGCAACGCTGCTTTGGGGGACCCCTGCC 541
77 yrProProProSerHisGlyHisLeuProProSerValGlyPro... 92
542 ACTCCACGGAGGT.....CCTGCCGC.....CCACTGCC 573
93 ProProHisArgGlyHisLeuProProSerArgGlyPheAsnProProPr 109
574 CGTCTCCTCC.CTGTGGCTCAGCATGGCGCGCTCACAACTCTGTAGC 622
109 oSerProValIleSerProSerHisProProSerTyrGlyAla...P 125
623 CCTGACCC.....GGGCCAGGCCAA...CAACAGTGAAGCAGCACCA 663
125 roProProSerHisGlyProGlyHisLeuProSerHisGlyGlnArgPro 141
664 TGACTACTGCTCCTGGTGGCT.....GCAGGAGGCAC..... 698
142 ProSerProSerHisGlyHisAlaProProSerGlyGlyHisThrProPr 158
699 .....CAGGCTCTCTCGCTGGACGGCGCTCTGGCGCTGTGG 741
158 oArgGlyGlnHisProProSerHisArgArgProSerProPro..... 172
742 CATCTCGCAGGAGC.....GCCTCTCGCTGCTGCT 770
173 ..SerArgHisGlyHisProProProProThrTyrAlaGlnProPro 188
771 CGCCCTTTCCTGTCAGCGCGCAGCTCAGCCACCTCGGCTCATG 820
189 ThrProIleTyrSerProSerProGlnValGlnProProProThrTyrSe 205
821 CCTGGGCACCTTCTGGCCCTTGGTGTGGCCAGGCTGACAGTGCACGCG 870
205 rPro.....ProProp 209
871 CTCCTGACCGCTCCCGCAGCTCGGCTCTGAGGGCCAGGCCAGGTGG 920
209 roThrHisValGlnPro..... 214
921 ACAGTGGCGGGCTCAGACACCGAGGCTCGGATGGGGCGGAGGGCTG 970
215 .....ThrProSerProSer.ArgGly..... 222
971 GCGGCGACGACCTCGGGGGCGGACCTGGGCGCTGCGGTAGCACTGCG 1020
223 ....HisGlnProGlnProProThrHisArgHisAlaProProThrHis 237
1021 CTGGTGGAGCACCATGCGCGCTCGCTTCGACGAGTGGGAAGTACAG 1070
238 ArgHisAlaProProThrHis..... 244
1071 CGGCCNAGCTGATGTGGCTGGCGGGCCAGCACCTTGCCTGACGGCCTT 1120
245 .....GlnPros 247
1121 GACCTGGCGCCCTCAAGCGCCAGCGAGGCTCTTCTGCTACTCGG 1170
247 eProLeuArgHisLeuProProSerProArgGlnProGlnProPro 263
1171 CCACTGGGACCAAACTGCAGCTACCTGCTGTGTACAGCCCGCA 1220
264 ThrTyrSerProProProProAlaTyrAlaGlnSerProGlnProSerPr 280

```

```

1221 ACGTGTGAAGGCTGCCCCCTGCTGCTTTGGGCTGGCGCCCAACACACA 1270
280 oThrTyrSerPro.ProProProThrTyrSerProProProProSerPro 296
1271 CTCAAAGTCACTGCGCGCCAGGCTGCGCTCTTGGTGTGGAAAGTGTAG 1320
297 IleTyrSerProProProProAla..... 304
1321 GCTGGTGCAGGCTGCTCCCACTGCTTCTTACTCCCTCCCTGCGGAGCCT 1370
305 ....TyrSerProSerProProProProThrProThrProThrPheSerProp 320
1371 CTTGCCCCCAAAAGTGGCTGCTGCTCTCTCCCTCTCTCCCTCCAC 1420
320 roProProAlaTyrSerProProProThrTyrSerProProProProThr 336
1421 CCACCTCACCTCCCTCC 1438
337 TyrLeuProLeuProSer 342
seq_name: p1r2:T43556
seq_documentation_block:
Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T43556
R:Zankel, T.C.; Ow, D.W.
submitted to the EMBL Data Library, December 1997
A:Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pombe
A:Reference number: 222575
A:Accession: T43556
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-574 <ZAN>
A:Cross-references: EMBL:AF038575; PIDN:AAB92587.1
A:Experimental source: strain JS21
C:Genetics:
A:Gene: wsp1
A:Map position: I
A:Introns: 72/3; 519/3; 564/1
alignment_scores:
Quality: 210.50 Length: 452
Ratio: 1.132 Gaps: 22
Percent Similarity: 41.150 Percent Identity: 25.664
alignment_block:
US-09-786-136-4 x T43556 ..
Align seg 1/1 to: T43556 from: 1 to: 574
234 CCAGAGCCCATGCCCCGAGGTGCTGCGCTTCAGCCCTTCAGACAAAGTAAG 283
159 ProSerAlaAlaProProValProGlyLysGluAsnTyrAsnAlaValG1 175
284 GTCAGCTCTGCCCTCCCTCTTACCTTGCCTGCTAGCTGCTGGATGCTAC 333
175 ySerLysSerProAsnGluProGluLeu..... 184
334 TACTAGGAGGTCTGCTGGGGCCCTGCAGGTGTGCAGCTCAGAGCCCG 383
184 ..... 184
384 CTGAGCCCCCAGGNAACCCCTCCCTGCTCTCACAGCCATCTAGATGACGT 433
185 LeuAsnSerLeuAspProSerLeuIleAspSerLeuMetLysMetGlyI1 201
434 CCTCTGCCCACTGTGTCTACTCTAAAGGACTTCAGAGAGGCT...CTCC 480
201 eSerGlnAspGlnIleAlaGluAsnAlaAspPheValLysAlaTyrLeuA 218

```

481 AGCAGGCGCTGGACTCGGACCAAAATGCAACTCCAAAGCGTCTTTGG 530
: : : : : ||||| ||| : : : : :
218 snGluSerAlaGlyThrProThrSerThrSerAlaPro... 230
531 GGGACCGCTGCTCCACTCCACAGGAAGTCTCTCGCGCGCCACCTCCCGCTCT 580
||| : : : : : ||| : : : : :
231ProThrProThrSerThrSerArgProPro... 242
581 CCCTGTGGGCTCAGACTGAGGCGCGCCGCTCACAACACTGTAGCCCTGACCC 630
242 242
631 GGGCCAGCCCAACACAGACTCAAGGACGACCATGACTACCTGCCCTTGG 680
||| : : : : : ||| : : : : :
243 ..GluArgValProSerValSerAlaProAlaProProProPro... 257
681 TGGCGGTGACGAGGACCAAGCTCTCTCTGCGCTGACGCGCTCTCTGTCG 730
258ProProSerAsnGlyThrValSerSerPr 267
731 GCGCGTGGCATCTCGCAGGAGCGCTCTGCGCTGCGCTGCGCTTGG.. 779
||| : : : : : ||||| : : : : :
267 oPro.....AsnSerProProArgProIleA 276
780CCGTGCACCGCGCCAGCTCAGCCCGCCCTGCG 812
||| : : : : : ||||| : : : : :
276 laProValSerMetAsnProAlaIleAsnSerThrSerLysProProLeu 292
813 CCTCATTCGCCCTGGGACTTCTGGCGCTTGGCCAGCGGTGACAGT 862
||| : : : : : ||| : : : : :
293 ProProProSerSerArgValSerAlaAlaLeuAla...AlaAsnLys 308
863 GCGACGCGCTCTGCGAGCGCGCTCGCCAGCTCGGCTCTGAGGGCGCAGG 912
||||| : : : : : ||||| : : : : :
308 sLysArgProProProProProProSerArgArgAsnArgGlyLysP 325
913 CCAGGTGGACAGTGGGGGGCTCAGACA..... 941
||| : : : : : ||| : : : : :
325 roProLleGlyAsnGlySerSerAsnSerLeuProProProProPro 341
942CCGAGGCGCTCCGATGGGGCG 961
||| : : : : : ||||| : : : : :
342 ProProArgSerAsnAlaAlaGlySerIleProLeuProProProGlnGlyAr 358
962 GAGGGC.....TGGGGCGCACCGACCTCGCGGGCGCG 993
||| : : : : : ||| : : : : :
358 gSerAlaProProProProProProProProProProSerThrGlyArgG 375
994 GACCTGGGCGCTCCGCTAGCAGCTCGCTGGCTGGAGCACCGATGCGCGG 1043
||| : : : : : ||| : : : : :
375 lnProProProLeu..SerSerSerArgAlaValSerAsnPro..... 388
1044 CTGCTTTCGACGAGTGGGAACTGACAGCGCGCCAAAGGCTGATGCTGCTG 1093
388 388
1094 GGGCCCGCAGACTGCTGAGCGCTGACCTGGCGCGCTCAAGCGCGC 1143
||| : : : : : ||| : : : : :
389ProAla.....ProProProAlaIlePro...GlyArg 398
1144 AGCCGAGGCGCTCTCTGCTACTGCGCCACTCGGA...CCAAACCTGC 1190
||| : : : : : ||| : : : : :
399 SerAlaProAlaLeuPro.....ProLeuGlyAsnAlaSerArgTh 412
1191 AGCTACACCTGCTGCTACAGCCGACGACCTGTGAAGCTGCGCCCT 1240
||| : : : : : ||| : : : : :
412 rSerThrProProVal..ProThrProProSer.....LeuProPro 425
1241 GCTGCTGGGTGGCGCCCAACCAACACTCAAGTCACTGCGCGCCAG 1290
||| : : : : : ||| : : : : :
426 SerAlaProProSerLeuProProSerAlaProProSerLeuProMetGl 442

1291 GGCTGGCGCTCTTGGTGGTGGGAAAGTGTAGGCTGTGCGCAGCGCTGCC 1340
||| : : : : : ||||| : : : : :
442 yAla.....ProAlaAlaProP 448
1341 CACTGCTTCTTACTCCCTCCCTGGAGCGCTCTTGGCCGCCACAAAAGTGC 1390
||| : : : : : ||| : : : : :
448 roLeu.....Pro.ProSerAlaProIleAlaProProLeuProAl 461
1391 TCGCGCTGCT 1440
||| : : : : : ||| : : : : :
461 aglyMetProAlaAlaProProLeuProProAlaProAlaProAlaPro 478
1441 CCTCTCAGCTCCCTCAACACAGTGGAGGGTAGAGCCACAGTCCCA 1490
||| : : : : : ||| : : : : :
478 roAlaProAlaProAla.....ProAlaAlaPro 487

seq_name: pir2:T38819

seq_documentation_block:

wiskott-aldrich syndrome protein homolog 1 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38819
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997

A:Reference number: 221813

A:Accession: T38819

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-574 <CON>

A:Cross-references: EMBL:Z98980; NID:el060691; PIDN:CAB11718.1; GSPDB:GN00066; SPDB:5

A:Experimental source: strain 972h-; cosmid C4F10

C:Genetics:

A:Gene: wsl1; SPDB:SPAC4F10.15c

A:Map position: 1

A:Introns: 72/3; 519/3; 564/1

alignment_scores:

Quality: 210.50 Length: 449

Ratio: 1.132 Gaps: 23

Percent Similarity: 41.425 Percent Identity: 27.171

alignment_block:

US-09-786-136-4 x T38819 ..

Align seg 1/1 to: T38819 from: 1 to: 574

234 CCACAGGCGCATGCCGAGGCTGCTGGCTTCGAGCCCTTCAGACAAAGTAAG 283

||| : : : : : ||| : : : : :
159 ProSerAlaAlaProProValProGlyLysGluAsnTyrAsnAlaValG1 175

284 GTCAGCTCTGCCCGCTCTCTGCTTCTACTTGCCTGTAGCTGTGGATGCTAC 333

||| : : : : : ||| : : : : :
175 ySerLysSerProAsnGluProGluLeu..... 184

334 TACTAGGAGGTCTGCTGCGGGGCGCTGCAGGTGTGCAGTCTCAGAGCCGG 383

184 184

384 CTGAGCGCCCGAGNACCCCTCTCTCTCAGAGCATCTAGATGACGT 433

||| : : : : : ||| : : : : :
185 LeuAsnSerLeuAspProSerLeuIleAspSerLeuMetLysMetGly1 201

434 CCTCTGCCACCTGTTGTCTACTCTAAAGGACTTCAGAGAGGCT...CTCC 480

: : : : : ||| : : : : :
201 eSerGlnAspGlnIleAlaGluAsnAlaAspPheValLysAlaTyrLeuA 218

481 AGCAGGCGCTGGGCTCGGACCAAAATGGCAACTCCAAAGCGTGTGG 530

: : : : : ||| : : : : :
218 snGluSerAlaGlyThrProThrSerThrSerAlaProProIle..... 232

531 GGGACCGCTGCCACTCCCGAGGAGGTCTCTCGCGCCCGCCACTCCCGCTCT 580

```

233 .....ProPro.SerIleProSerSerArgProProGluArgValProS 247
581 CCTGTGCGCTCAGCATGGCGCGCGCTCACAAACTCTGTAGCCCTGACCC 630
||||| ||||| |||||
247 erLeuSerAlaProAlaProProIle.....Pro 257
631 GGGCCAGGCCACACACAGTGAAGGAGGAGCCACCATGACTACCTGCCCTGG 680
|||||
258 ProPro..... 259
681 TGGGGCTGCAGGAGCACCAGGCTCTTCGCCCTGGAGC.....CGCC 723
260 .....SerAsnGlyThrValSerSerProProAsnSerProProArgP 274
724 CTTCTGCGCGCTGTGCGCATCTCGCAGAGCGCCTCTGCTGCGCTGCTGCG 773
|| |||||:
274 roIleAlaProValSerMetAsn..... 281
774 CTTTTCCTGTCACCGCGCCAGCCTCAGCCCGCCACCTCGGCTCATTTGCC 823
|||||: ||||| ||||| |||||
282 .....ProAlaIleAsnSerThrSerLysProProLeuProProSe 296
824 TGGGCATCTTCTGGCCCTGCTGTGGCCAGGCTGACAGTGCCACGCGCTC 873
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
296 rSerArgValSerAlaAlaAlaLeuAla...AlaAsnLysLysArgProp 312
874 CTGACGCCGCTCCCGAGCTCGGGCTCTGAGGGCCAGGCCAGGTGGACA 923
|| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
312 roProProProProProSerArgArgAsnArgGlyLysProProIleGly 328
924 GTGGGGGGGGCTCACACA..... 941
329 AsnGlySerSerAsnSerSerLeuProProProProProProArgSe 345
942 .....CCGAGCGCTCCGATGGCGGAGGCG..... 968
345 rAsnAlaAlaGlySerIleProLeuProProGlnGlyArgSerAlaProp 362
969 .....TGGGGGCGCAGCATCTGGGGCGCGGCGGACCTGGGCGCA 1004
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
362 roProProProProProArgSerAlaProSerThrGlyArgGlnProPro 378
1005 CTGCGCTACCATCGCTGGCTGGAGCAGCCGATGGCGCGCTGCTGCTGAC 1054
|| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
379 Leu.SerSerSerArgAlaValSerAsnPro..... 388
1055 GAGTGGGAACCTGACAGCGGCGCAAGGCTGATTGCTGGCTGCGGGGCCAGCA 1104
|| |||||
389 .....ProAla 390
1105 CTTGCTGAGCGCTTGACCTGGCGCCCTCAAGCCCGCAGCCGAGGCG 1154
|| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
391 .....ProProProAlaIlePro...GlyArgSerAlaProAl 402
1155 TCTTCTGCTACTGCGCCACTGGGA...CCAAACCTGCAGCTACACCTG 1201
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 aleuPro.....ProLeuGlyAsnAlaSerArgThrSerThProp 416
1202 CTGTGTACAGCCAGCGAACCTGTGAAGGCTGCCCGCTGCTGCTGGGC 1251
|| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
416 roVal.ProThrProProSer.....LeuProProSerAlaProPr 429
1252 TGGGGCCCGCCACACACTCAAGTCACTGCGCGCCAGGCTGGCCCTCT 1301
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
429 oSerLeuProProSerAlaProProSerLeuProMetGlyAla..... 443
1302 TGGTGTGGGAAGTGTAGGCTGGTGGCAGCGCTGCCCCACTGCTTCTT 1351
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
444 .....ProAlaAlaProProLeu..... 449
1352 ACTCCCTCCCTGGAGCCCTCTTGGCCCCCACAAGAGTGCCTGCCCTGCT 1401
|| |||| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

```

```

450 ...Pro.ProSerAlaProIleAlaProProLeuProAlaGlyMetProA 465
1402 CTCCTCCCTCTCTCCACCCACCTCACACTCCCTCCATCTCTGAGCTC 1451
: ||||| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
465 laAlaProProLeuProProAlaAlaProAlaProProAlaProAla 481
1452 CTTGCACACAGTGGAGGGTAGAGAGCCACAGTCCCA 1490
||||| ||||| ||||| |||||
482 ProAla.....ProAlaAlaPro 487

seq_name: pir2:T18535
seq_documentation_block:
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch
A:Reference number: Z18955; MUID:9803440
A:Accession: T18535
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SH1>
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

alignment_scores:
Quality: 210.50 Length: 509
Ratio: 0.873 Gaps: 24
Percent Similarity: 47.348 Percent Identity: 25.147

alignment_block:
US-09-786-136-4 x T18535 ..
Align seg 1/1 to: T18535 from: 1 to: 1151

202 GTGTGGGGGCTCCACAGACTGCTCTGTGGTTCCACAGGCGCATGCCGGA 251
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
188 ValThrAlaAlaSerGlnSerAlaProLysAlaAlaThrAspAlaAla 204
252 GGTCTGGCTTCGAGC.....C 268
|| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
204 aValThrAlaAlaSerGlnSerAlaProLysAlaAlaThrValGluValLysP 221
269 CTTCAGACAGTAAGTCAGCTGCGCCCTCTGCTTCTGCTTCTGCTTCTGCT 318
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
221 roAlaAlaAlaAlaAlaLysGluAlaLysAlaValThrAlaAla 237
319 AGCTGTGGATGCTACTAGGAGGTCCTGCTGGGGCGCTGCGAG... 364
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
237 aAlaAlaProLysAlaThrAlaGluAlaLysProAlaProValThrSerP 254
365 .....GTGTGAGCTCAGAGCCCTCAGGCC.....CCAGGAACC 400
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
254 roThrIleProCysSerSerAlaGluAlaLysProLeuThrAlaAlaSer 270
401 CCTCCTGCCTCTCACAGCCTCTAGATGTCAGCTCCTCTGCCC..... 442
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
271 ProThrAlaSerLysAlaThrAlaGluAlaLysProValProAlaThrAl 287
443 .ACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGGCGCCT 491
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
287 aserLeuMetAlaThrLysValThrAlaGluAlaLysProAlaProSerP 304
492 GGGACTCGGACCAAAATGGCAACTCCAGCGTGTGTTGGGGGACCCCTGCC 541
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
304 roSerValProLysAlaThrThrAspThrLysAlaValThrAlaThrAla 320
542 ACTCCCGAGGAAGTCTCT..... 559
|| |||||
321 ...ProLysAlaGlyProAspValLysProAlaValAlaValCysAlaGI 336

```



```
560 .....CGCGCCACCTCCCGCTCTCCCTCGGCTCAGCAGTGGCGCC 605
    :::: ||||| ||||| :::: |||||
336 uAlaLysProAlaProProProGlnGlnLeuProLys.AlaAla 352
    :::: ||||| ||||| :::: |||||
606 GTCACAACTCTCTAGCCCTGACCGCGGCAGGCCAACACAGTGAAGGC 655
    :::: || ||||| :::: |||||
353 AlaAla.....AlaAlaProThrGlyThroLule 362
    :::: ||||| ||||| :::: |||||
656 AGGCACATACATACCTGCGCTTGGTGGGCTCGCAGGAGCCAGGCTC 705
    :::: ||||| ||||| :::: |||||
362 uLysProAlaAlaProHisGlySerProArgAlaAsnSerHist 379
    :::: ||||| ||||| :::: |||||
706 CTTCCGCTGGAGCGCTCTCGCGCTGCGCATCTCCAGG... 752
    :::: ||||| ||||| :::: |||||
379 hrValThrValThrProProAsnValProArgAlaAlaAlaThrVal 395
    :::: ||||| ||||| :::: |||||
753 ...AGCGCTCTCGCTGCTCGCCCTTTGCCGTCAGCGCGCCAGCCTC 799
    :::: ||||| ||||| :::: |||||
396 ProThrAlaGlyAlaValProLysAlaSerThrGlyThrProAlaAl 412
    :::: ||||| ||||| :::: |||||
800 AGGC.....CCACCTCGGC 813
    :::: ||||| ||||| :::: |||||
412 aAlaProGlnGlnProValProLysAlaAlaProValThrProSerP 429
    :::: ||||| ||||| :::: |||||
814 CTCATTGCCCTGGCCTCTCGGCCCTGCTTGGCCAGGCTGACAGTG 863
    :::: ||||| ||||| :::: |||||
429 roGlnGlnAlaValProArgAlaAlaThrAlaAlaAlaProValThr 445
    :::: ||||| ||||| :::: |||||
864 CCACGGCTCTCTGCA.....GCCGCTCCCC 889
    :::: ||||| ||||| :::: |||||
446 ProGlnGlnProValThrLysAlaAlaThrThrAsnAlaThrProPr 462
    :::: ||||| ||||| :::: |||||
890 AGTCGGGCTCTGAGGCCAGCCAGGTGGAGTGGCGGGGCTCAGA 939
    :::: ||||| ||||| :::: |||||
462 oProGlnProileProLysAlaAlaThrThrAlaThrProValT 479
    :::: ||||| ||||| :::: |||||
940 CACGAGCGCTCGATGGCGCGGAAGGCTGG...CGCGCAGCAGCTGC 986
    :::: ||||| ||||| :::: |||||
479 hrProGlnGlnProileProLysAlaGlyThrAspAlaAlaProPro 495
    :::: ||||| ||||| :::: |||||
987 GGGCGCGAGCTGGCGCTGCGCTGACCTCGCTGGAGCACCGA 1036
    :::: ||||| ||||| :::: |||||
496 AlaValProLysAlaPro.....SerAspGlyArgAlaAlaTh 508
    :::: ||||| ||||| :::: |||||
1037 TGGCCCTGCTTGCAGAGTGGGAGTGCAGAGCGGCCAGGCTGATTG 1086
    :::: ||||| ||||| :::: |||||
508 rProGlyValProAsnAlaAlaThrAspProGlnLysProProThr. 524
    :::: ||||| ||||| :::: |||||
1087 CTGCTCGGGCGCAGCCTGCTGACGGCTTGACCTGGCGCCCTCA 1136
    :::: ||||| ||||| :::: |||||
525 .....ProGlnSerValProSerAlaValThrGluPro..... 535
    :::: ||||| ||||| :::: |||||
1137 AGCGCGAGCCGAGGCTCTCTGCTACTGCGCCACTGGGACCAAAAC 1186
    :::: ||||| ||||| :::: |||||
536 LysProGln.ProArgAlaAlaProPro.....ProSerA 547
    :::: ||||| ||||| :::: |||||
1187 CTGAGCTACACCTGCTGCTACAGCCAGCAAGCTGTGAAGGTGCC 1236
    :::: ||||| ||||| :::: |||||
547 snGluAlaThrProAlaValProSerProSerProAsnLeuLysSer..P 563
    :::: ||||| ||||| :::: |||||
1237 CCCTG.....CTGTTGGCTGGCGCCGCC 1262
    :::: ||||| ||||| :::: |||||
563 roLeuProThrIleProLysProValProLeuMetAlaLeuThrProGln 579
    :::: ||||| ||||| :::: |||||
1263 CCA.....ACACACTCAAGTCACTGCGCGCCAGGGCT 1294
    :::: ||||| ||||| :::: |||||
580 ProValThrAlaGlnMetValThrGlnLeuAlaAlaThrLysPro..... 594
    :::: ||||| ||||| :::: |||||
1295 GGCCTCTGTGTGGGAAAGTGTAGCTGGTCCAGCCTGTC..... 1337
    :::: ||||| ||||| :::: |||||
595 .....SerProIleValProL 600
    :::: ||||| ||||| :::: |||||
```

```
1338 .....CCCACTGCTTCTACTCCCTCGAGCCCTTGTGCCCC 1379
    :::: ||||| ||||| :::: |||||
600 yAlaSerProLysAlaLeuMetThrProProProProGlyLeu 616
    :::: ||||| ||||| :::: |||||
1380 ACAAAAGTGCCTGCTGCTCTCTCCCTCTCCACCCCACTCACA 1429
    :::: ||||| ||||| :::: |||||
617 ProArgAlaLeuAlaAlaLysLeuGlyLeuPro.....Se 630
    :::: ||||| ||||| :::: |||||
1430 CTCCTCTCATCTCTGAGTCCCTGCAACAGTGGAGGCTAGAGGC 1479
    :::: ||||| ||||| :::: |||||
630 rSerProValAlaSerAlaMetHisAlaLysValThrProArgProLeuP 647
    :::: ||||| ||||| :::: |||||
1480 .CACAGTCCCAATCTATG 1499
    :::: ||||| ||||| :::: |||||
647 roAlaSerProValProMet 653
```

seq_name: pir2:I78557

seq_documentation_block:

N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000

C:Accession: I78557; I58158; D45219

R:Monyer, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.

Neuron 12, 529-540, 1994

A:Title: Developmental and regional expression in the rat brain and functional proper

A:Reference number: I58158; MUID:94206533

A:Accession: I78557

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1323 <RES>

A:CROSS-references: GB:L31612; NID:G469068; PIDN:AAC37647.1; PID:G469069

A:Accession: I58158

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-66,'V',68-1323 <RE2>

A:CROSS-references: GB:L31611; NID:G469066; PIDN:AAC37646.1; PID:G469067

R:Ishii, T.; Moriyoshi, K.; Sugihara, H.; Sakurada, K.; Radotani, H.; Yokoi, M.; Akaz

J. Biol. Chem. 268, 2836-2843, 1993

A:Title: Molecular characterization of the family of the N-methyl-D-aspartate recepto

A:Reference number: A45219; MUID:93155102

A:Accession: D45219

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1265-1323 <ISH>

A:Experimental source: brain

A:Note: sequence extracted from NCBI backbone (NCBIP:124265)

C:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology

F:451-879/Domain: glutamate receptor homology <GRH>

alignment_scores:

Quality: 210.50 Length: 441

Ratio: 1.341 Gaps: 22

Percent Similarity: 35.601 Percent Identity: 24.717

alignment_block:

US-09-786-136-4 x I78557 ..

Align seg 1/1 to: I78557 from: 1 to: 1323

353 GGGGCCCTCAGTGTGTCAGCTCAGCCCGCTGAGCCCGCCAGGACCC 402

||| :::: |||||:|||| ||| |||||:|

889 GlyMetTyrSerCysSerAlaGluAlaAlaProProProAlaLysPr 905

403 TCCTGCTCTCACGCCATCTAGATGTCAGCTCTCTGCTGCTGTGTCT 452

||| ||||| |||||:|||| ||| |||||:|

905 oProPro.....ProGlnProLeuProSerProAlaL 917

453 AC..... 454

|||

917 yrProAlaAlaArgProProProGlyProAlaProPheValProArgGlu 933


```

||||:|||| ||| |||||||||| ||| :||
479 GlySerProProAlaThrSerProSer...ProGlnValGlnProPro 494
600 GCGCGCTGACAACTCTGTAGCCCTGACCCGGCCAGGCCAACACAGT 649
||||| ||| :|||:||||: ||| :|
495 AlaAlaSerThrProProProSerLeuValLysLeuSerProProGlnAl 511
650 GAAGGAGGAGGACC.....ATGACTACTGCGCCTTGTGCG 684
: :|||:||||| ||||| |||
511 aProValGlySerProProProValLysThrThrSerPro..... 525
685 GCTGAGGAGGACCAAGCTCTTCGCGCTGGACGGCCCTTGTGCGCG 734
||||| :||| ||| |||
526 .....ProAlaProIleGlySerProSerProProProPro 537
735 CTGTGGCCATCTCGCAGGAGCGCTCTGCGCTCGCTCGCCCTTGTG 779
:| :|||:||||: ||| |||:|
538 Val.....SerValValSerProProProValLysSer 549
780 ....CCGTGACCGCGCCAGCCCTCAGCCACCTCGGCCTCAT..... 818
||| |||||:||||:||||| |||:|
549 rProProProProAlaProValGlySerProProProGluLysSerP 566
819 .....TGCCCTGGGCATCTTGGGCCCTGGTGTGGCCAGGTGACATG 863
||| ||| ||| |||:|
566 roProProPro.....AlaProValAlaSer 574
864 CCACGGCTCTGACGCGCTGCCAGCTGCCGCTCGAGGCTCGATGGCG 913
|||||:||||| ||||| :|||
575 ProProProProValLysSerProProProProThrLeu..... 587
914 CAGGTGGACATGGGGGGCTCAGACACGAGGCTCGATGGCGCGGA 963
:| :|||:||||:||||| :|||
588 .....ValAlaSerProProProProVal..... 595
964 AGGCTGGGGGACCGACCTGCGGGCGGACCTGGCCACTGCGGTAG 1013
:| :||| ||| |||
596 .....LysSerProProProProAlaProVal 604
1014 CACTGCTGCTGGAGCAGCGCTGCGCTGCTGCTGCGAGTGGGA 1063
||||| :||| ||| |||
605 AlaSerProProProProValLysSerProProProProValAl 621
1064 CTGACAGCGGCAAGCTGATGCTGGCTGGCGGCCAGCACTGCTGTA 1113
||| :|||:||||:|||||
621 aSerProProProProAla..ProValAlaSerSerProProMetLys 637
1114 CGGCTTGACCTGGCGCTCAAGCGCGCAGCGCGGCTTCTCTGTC 1163
||| ||| ||| :||| |||
638 SerProProProProThrProValSerSerProProPro.....ProGl 652
1164 TACTGCGCCACTGGGACCAAAACCTGCA...GCTACACCTGCTGTGCTAC 1210
:|||| ||| ||||| :||| |||
652 uLysSerProProProProProProAlaLysSerThrPro..... 665
1211 AGCCAGCAAGCTGTAAGCTGCGCCCTGCTGCTGCTGGCTGGCGCCC 1260
||| ||||| :|||:||||: |||
666 ..ProProGluGluTyProThr..ProProThrSerValLysSerPr 681
1261 ACCCAACACACTCACTGCTGCGCGCCAGGCTGGCCCTTGTGCTGG 1310
||||| ||||| ||||| |||
681 oProProGlu...LysSerLeuProProProThr..... 691
1311 GAAAGTGTAGGTGGTCCAGCGCTGTCGCCACTG..... 1345
|||:||||: |||||
692 .....LeuIleProSerProProProGlnGluLysProThrPro 704
1346 CTCTTACTCCCTCCCTGGAGCCCTTGTGCCCCCAAAAGTGC..... 1390
||||||| ||||| |||||
705 ProSerThrProSerLysProProSerSerProGluLysProSerProPr 721
1390 ..... 1390

```

```

721 oLysGluProValSerSerProProGlnThrProLysSerSerProProp 738
1391 ..CTGCTGTGCTCTCTCCCTCTCCCTCCACCCACCTCACACTCCCTCC 1438
||||| ||||| ||||| |||||:|||||
738 roAlaProValSerSerPro...ProProThrProValSerSerProPr 753
1439 A..... 1439
753 oAlaLeuAlaProValSerSerProProSerValLysSerSerProProp 770
1440 .....TCCTCTGAGCTCCCTCTCAACACAGTGAAGGGTAGAGAG 1478
||||| ||||| ||||| |||
770 roAlaProLeuSerSerProProProAlaProGlnValLys..... 783
1479 CCACAGTCCCAATCTATGCAATAAAGTGCA 1511
:||||| ||||| |||||
784 ...SerSerProProProValGlnValSerSer 793

```

seq_name: p1r2:S27224

seq_documentation_block:

N-methyl-D-aspartate receptor epsilon-4 chain - mouse

C.Species: Mus musculus (house mouse)

C.Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999

C.Accession: S27224

R.Ikeda, K.; Nagasawa, M.; Mori, H.; Araki, K.; Sakimura, K.; Inoue, Y.

FBS Lett. 313, 34-38, 1992

A.Title: Cloning and expression of the epsilon-4 subunit of the NMDA receptor channel

A.Reference number: S27224; MUID:93050214

A.Accession: S27224

A.Status: preliminary

A.Molecule type: mRNA

A.Residues: 1-1323 <INEX>

A.Cross-references: EMBL:D12822

C.Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology

C.Keywords: transmembrane protein

F:451-879/Domain: glutamate receptor homology <GRI>

alignment_scores:

Quality: 205.50 Length: 443

Ratio: 1.309 Gaps: 26

Percent Similarity: 35.440 Percent Identity: 26.185

alignment_block:

US-09-786-136-4 x S27224 ..

Align seg 1/1 to: S27224 from: 1 to: 1323

```

353 GGGGCCCTGCAGGTGTGACGTGTCAGAGCCGCTGAGCCGCCAGGAACCCC 402
||| :|||:||||:||||| ||| |||||:|
889 GlyMetTySerCysCysSerAlaGluAlaAlaProProAlaLysPr 905
403 TCCTGCTCTCACAGCCATCTAGATGCACTCTCTGCTGCCACTGTTCT 452
||||| :|||:||||:|||||
905 oProPro.....ProProGlnProLeuProSerProAlat 917
453 ACTCTAAAGGACTTCACAGAGGCTCTCCAGAGCGCGCTGGGACTCGGAC 502
||| ||| |||||
917 yr.....ProAla..... 919
503 CAAATGGCAACTCAAGCGTGTGTTGGGGACCCCTGCCACTCCACGGA 552
||| |||
920 .....AlaArgProPr 924
553 AGTCTCTCGCGCCACCTCCCGTCTCTCTGCTGCGCTCAGCATGGCC 602
||||| ||||| :|||
924 oGlyProAlaProPheValProArgGluArgAlaAlaAlaAspArgTrp 941
603 GCGCTCACAAACTCTGTAGCCCTGACCCGGCCAGGCGCAACACAGTGA 652
||||| ||| :|||:||||| |||
941 rgArgAlaLysGlyThrGlyPro...ProGlyGlyAlaAlaLeuAlaAsp 956

```


OM of: US-09-786-136-4 to: SwissProt_40:* out_format : pfs
 Date: Sep 19, 2002 5:46 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+2np.model -DEV=xlh
 -O/cgn2_1/USPTO.spool/US09786136/runat_16092002_141245_19278/app_query.fasta_1.1592
 -DB=SwissProt_40 -QFMT=fastan -SUFFIX=rs -GAPOPT=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum2
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=LOCAL -THR_SCORE=pct
 -R_MAX=100 -THR_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs
 RM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
 USER=US09786136_CGNL_1.22 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-786-136-4
 Query length: 1511
 Database: SwissProt_40
 Database sequences: 105724
 Database length: 38719550
 Search time (sec): 37.740000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_40:T2D3_HUMAN	240.50	213.81	3.0e-05	1083	I 000268 homo sapiens (human)
SwissProt_40:GPI_CHLRE	235.00	213.62	6.0e-05	555	I Q9fpg6 chlamydomonas reinhardtii (chlamydomonas)
SwissProt_40:YH1_EBV	218.50	198.05	0.0004	660	I P03181 epstein-barr virus (strain 660)
SwissProt_40:EXTN_TOBAC	214.00	194.57	0.0006	620	I P13983 nicotiana tabacum (cigarette)
SwissProt_40:NME4_MOUSE	211.50	187.18	0.0007	1323	I Q03391 mus musculus (mouse)
SwissProt_40:NME4_MOUSE	210.50	186.30	0.0008	1323	I Q62645 rattus norvegicus (rat)
SwissProt_40:NME4_MOUSE	205.00	186.50	0.0017	641	I P03211 epstein-barr virus (strain 641)
SwissProt_40:NME4_MOUSE	200.00	177.09	0.0027	1336	I Q15399 homo sapiens (human)
SwissProt_40:TEGU_EBV	199.50	170.76	0.0026	3149	I P03186 epstein-barr virus (strain 3149)
SwissProt_40:TEGU_EBV	196.50	178.89	0.0043	660	I P03181 epstein-barr virus (strain 660)
SwissProt_40:MUC2_HUMAN	196.50	164.73	0.0034	5179	I Q02817 homo sapiens (human)
SwissProt_40:YVNA_PRYKA	192.00	168.34	0.0064	1733	I P33485 pseudorabies virus (dog)
SwissProt_40:YVNA_PRYKA	191.00	167.34	0.0085	415	I P02732 canis familiaris (dog)
SwissProt_40:CA12_MOUSE	189.50	167.34	0.0086	1459	I P28481 mus musculus (mouse)
SwissProt_40:CA12_MOUSE	189.50	162.51	0.0080	2944	I Q02388 homo sapiens (human)
SwissProt_40:CA12_MOUSE	188.50	166.67	0.0097	1418	I P02458 homo sapiens (human)
SwissProt_40:SYNL_HUMAN	186.00	169.30	0.0139	705	I P17600 homo sapiens (human)
SwissProt_40:SYNL_HUMAN	182.00	168.62	0.0228	489	I P27918 homo sapiens (human)
SwissProt_40:PROP_HUMAN	182.00	154.73	0.0180	3530	I Q9ukn7 homo sapiens (human)
SwissProt_40:SYNL_HUMAN	181.50	165.37	0.0230	706	I P17599 bos taurus (bovine)
SwissProt_40:VGLG_HSV2H	181.00	165.00	0.0243	699	I P13290 herpes simplex virus (herpes simplex virus)
SwissProt_40:WALP_HUMAN	180.50	166.83	0.0268	503	I Q43516 homo sapiens (human)
SwissProt_40:SYNL_RAT	180.50	164.52	0.0257	704	I P09951 rattus norvegicus (rat)
SwissProt_40:HFCL_HUMAN	180.00	156.78	0.0240	2035	I P51610 homo sapiens (human)
SwissProt_40:SE5_RAT	179.00	162.12	0.0298	825	I Q63003 rattus norvegicus (rat)
SwissProt_40:NTC3_MOUSE	179.00	155.01	0.0264	2318	I Q61982 mus musculus (mouse)
SwissProt_40:CA21_CHICK	178.50	158.24	0.0298	1362	I P02467 gallus gallus (chicken)
SwissProt_40:WAS1_HUMAN	177.50	163.49	0.0370	559	I Q92558 homo sapiens (human)
SwissProt_40:SYNL_MOUSE	177.50	162.24	0.0362	670	I Q88935 mus musculus (mouse)
SwissProt_40:PRP3_MOUSE	177.00	167.43	0.0421	296	I P05143 mus musculus (mouse)
SwissProt_40:SPD1_MOUSE	177.00	161.06	0.0368	747	I P19837 nephila clavipes (orb-weaver)
SwissProt_40:T2D3_MOUSE	177.00	158.51	0.0371	1083	I Q00268 homo sapiens (human)
SwissProt_40:DRPL_RAT	176.50	157.46	0.0378	1183	I P54258 rattus norvegicus (rat)
SwissProt_40:CA11_MOUSE	175.50	155.18	0.0413	1453	I P11087 mus musculus (mouse)
SwissProt_40:CA13_MOUSE	174.50	154.26	0.0461	1464	I P08121 mus musculus (mouse)
SwissProt_40:TEGU_HSV1	174.50	148.95	0.0421	3164	I P10220 herpes simplex virus (herpes simplex virus)
SwissProt_40:CA12_BOVIN	174.00	159.19	0.0535	671	I P02454 rattus norvegicus (rat)
SwissProt_40:CA12_BOVIN	174.00	158.45	0.0528	747	I P02459 bos taurus (bovine)
SwissProt_40:GSRL_HUMAN	174.00	153.61	0.0486	1509	I Q9nzm4 homo sapiens (human)
SwissProt_40:CA21_BOVIN	173.50	153.87	0.0520	1364	I P02465 bos taurus (bovine)

SwissProt_40:CALL_CHICK	173.50	153.44	0.0516	1453	I P02457 gallus gallus (chicken)
SwissProt_40:APXL_HUMAN	173.50	152.71	0.0510	1616	I Q13796 homo sapiens (human)
SwissProt_40:CA21_HUMAN	171.00	151.68	0.0687	1366	I P08123 homo sapiens (human)
SwissProt_40:CALL_CANFA	170.50	150.79	0.0721	1460	I Q9xsj7 canis familiaris
SwissProt_40:CA21_ONCMA	170.00	150.86	0.0769	1356	I Q93484 oncorhynchus mykiss

seq_name: SwissProt_40:T2D3_HUMAN

seq_documentation_block:

ID	T2D3_HUMAN	STANDARD;	PRT;	1083	AA.
AC	000268;	Q99721;	Q9BX42;	Q9BR40;	
DC	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DE	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	Transcription Initiation factor TFIID 135 kDa subunit (TAFII-135)				
DE	(TAFII135) (TAFII-130) (TAFII130)				
GN	TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=97336072; PubMed=9192867;				
RA	Mengus G., May M., Carre L., Chambon P., Davidson I.;				
RA	*Human TAF(II)135 potentiates transcriptional activation by the AP-2s				
RA	of the retinoic acid, vitamin D3, and thyroid hormone receptors in				
RA	mammalian cells.;				
RL	Genes Dev. 11:1381-1395(1997).				

SEQUENCE FROM N.A.

DELOUKAS P., MATTHEWS L.H., ASHURST J., BURTON J., GILBERT J.G.R., JONES M., STAVRIDES G., ALMEIDA J.P., BABAGE A.K., BAGULEY C.L., BAILEY J., BARLOW K.F., BATES K.N., BEARD L.M., BEARE D.M., BEASLEY O.P., BIRD C.P., BLAKE S.E., BRIDGEMAN A.M., BROWN A.J., BUCK D., BURRILL W., BUTLER A.P., CORDER C., CARTER N.P., CHAPMAN J.C., CLAMP M., CLARK G., CLARK L.N., CLARK S.Y., CLEE C.M., CLEGG S., COBLEY V.E., COLLIER R.E., CONNOR R., CORBY N.R., COULSON A., COVILLE G.J., DEADMAN R., DHAMI P., DUNN M., ELLINGTON A.G., FRANKLAND J.A., FRASER A., FRENCH L., GARNER P., GRAHAM D.V., GRIFFITHS C., GRIFFITHS M.N.D., GWILLIAM R., HALL R.E., HAMMOND S., HARLEY J.R., HEATH P.D., HO S., HOLDEN J.L., HOWDEN P.J., HUCKLE E., HUNT A.R., HUNT S.E., JEDOSCH K., JOHNSON C.M., JOHNSON D., KAY M.P., KIMBERLEY A.M., KING A., KNIGHTS A., LAIRD G.K., LAWLER S., LEIVISALHO M.H., LEVERSHA M., LLOYD C., LLOYD D.M., MCCORMACK L., MARSH V.L., MARTIN S.L., MCCONNACHIE L.J., MCLEAY K., MCMURRAY A.A., MILNE S., MISTRY D., MOORE M.J.F., MULLIKIN J.C., NICKERSON T., OLIVER K., PARKER A., PATEL R., PEARCE T.A.V., PECK A.I., PHILLIMORE B.J.C.T., PRATHALLINGAM S.R., PLUMB R.W., RAMSAY H., RICE C.M., ROSS M.T., SCOTT C.E., SEHRA H.K., SHOWNKEEN R., SIMS S., SKUCE C.D., SMITH M.L., SODERLUND C., STEWARD C.A., SULSTON J.E., SWANN M., SYCAMORE N., TAYLOR R., TEE L., THOMAS D.W., THORPE A., TRACEY A., TROMANS A.C., VAUDIN M., WALL M., WALLIS J.M., WHITEHEAD S.L., WHITTAKER P., WILLEY D.L., WILLIAMS L., WILLIAMS S.A., WILMING L., WRAY P.W., HUBBARD T., DURBIN R.M., BENTLEY D.R., BECK S., ROGERS J.;

*The DNA sequence and comparative analysis of human chromosome 20.;

Nature 414:865-871(2001).

[3]

SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.

MEDLINE=97098442; PubMed=8942982;

Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;

*Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100.;

Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).

-1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.

-1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).

-1- SUBCELLULAR LOCATION: Nuclear.

-1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.


```

217 .....ProProSerProPro.SerProAlaProProSerProProSe 230
1164 TACTGGCCCACTGGGACCAAAACCTGCAGCTACACCTGTGTGTCTACAGC 1213
      ||||| :|||:||||| ||| |||: :|||
230 :ProAlaProProSerProSerProAlaProProSerProValProp 247
1214 CAGCCCAACGTGTGAAGGCTGCCCTGTGCTGTGGCTGGCCGCCACC 1263
      ||||| ||| |||||:|:|
247 roSer .....ProAla.ProProSerPro.....AlaProPr 257
1264 CAACACACTCAAGTCACTGCGCCCGCAGGGCTGCGCTTGTGGTCTGGAA 1313
      |:|:|
257 oSerProLysProProAlaProPro..... 265
1314 AGTGTAGGTGGTGGCAGCGCTGTCGCCCACTGTCTTACTCCTCCCTG 1363
266 .....ProPro.....ProSerPro 270
1364 GAGCCCTCTTGGCCCAACAAAAGTGCCTGCTGCTCTCTCCCTCTCC 1413
      ||| ||| :|:|:|:|:|:|:|:|:|
271 ProProProProProProArgProProPheProAlaAsnThrProMetPr 287
414 TCCACCCCACTCACACTCCCTCCCTCTCTGAGCTCCCTGCAACACAG 1463
      ||||:|:| ||| ||||| ||| |||
287 oProSerPro.Pro...SerProProProSerProAlaProProThrPro 302
1464 TGAAGGGTAGAGAGCACAGTCCCAATCTATG 1499
303 ProThrProProSerProSerProVal 314
seq_name: SwissProt_40:YHL1_EBV

```

seq_documentation_block:

```

ID YHL1_EBV STANDARD; PRT; 660 AA.
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLFI protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; -: NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0BE3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

```

alignment_scores:

Quality: 218.50 Length: 586

```

Ratio: 0.967 Gaps: 34
Percent Similarity: 38.567 Percent Identity: 26.792
alignment_block:
US-09-786-136-4/rev x YHL1_EBV ..
Align seg 1/1 to: YHL1_EBV from: 1 to: 660
1481 TGGCTCTCTACCTCCATCGTGTTCAGGAGGAGCTCAGAGGATGGAGGG 1432
      ||| :|:| ||| ||| |||
148 TrpSerAlaArgAsnProGlyCysProArgThr..... 158
1431 AGTGTAGTGGGTGGGAGGAGGAGGAGGAGACAG..... 1395
      |||:|:|:|:|:|:|:|
159 .....TrpArgArgSerGlyAlaGlnArgGlyHisProp 171
1394 .....GCAGGCACCTTTTGTGGGGCAAGAG 1368
      |||
171 roProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla 187
1367 GTCCAGGAGGAGGTAAAGAGCAGTGGGGACAGGCTGGCACC..... 1323
      ||||| ||| ||| |||
188 AlaProGly.....AlaProGlyThrProAlaAlaProGlyPr 200
1323 ..... 1323
200 oGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArg 217
1322 .....GCCTACACTTTTCCAGCACCACAGA 1299
      ||| :|:|:|
217 LySerGlyProAlaAspProProAlaAlaAlaArgLeuProGluArg 233
1298 GGCAGCGCCCTGG.....GCGGCAGTGCAGTGTGAGTGTGTTGGT 1261
      :|:|:| |||:|:|:| ||| :|
234 GlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAl 250
1260 GGGCGCCCAAGCAGCAGGAGGAGGAGGAGC..... 1230
      :|:| ||| :|:|:|:|:|:|
250 aGlyProProThrArgSerGlyAlaAlaGlnArgThrHisArgA 267
1229 .....TTACACAGTTCGCTGGGCTGT..... 1209
      |||:|:|:|:|:|:|
267 rgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 283
1208 .....AGCAGCAGAGTGTAGTGCAGTGTTCCTCCAGTGGCGCAGT 1165
      :|:|:|:|:| ||| ||| |||||
284 TrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAla... 299
1164 AGCAGGAGAGCCCTCGGCTGGGCTTGAAGGAGCGGCGCAGGTCAAGCC 1115
      ||| ||||| ||| ||| |||
300 .....GlyGlnArgProSerGly...ProThrGlyGlyA 310
1114 GTCAGGCAAGTGTGGG.....CCGCGCAGCAGCAATCAGCCT 1077
      ||| ||||| ||| |||||
310 rgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 326
1076 TGGCCCGCTGCACTTCCACTCGTTCGAAGGAGCGGCGCATCGGTGCT.. 1029
      :|:|:|:|:| |||||
327 GlyGlyAlaAlaValProSer.....GlyAlaTh 336
1028 .....CCAGCCAGGCGAGTGTACGGCAGTGCCTCCGCGGCGCCCG 986
      ||| ||| |||||
336 rProHisProGluArg.....GlySerGlyProA 346
985 CAGGTGCGTGGCGCCAGCC...CTTCGCGCCCATCGGAGG..... 948
      || ||| |||||:|:|:|:|:|
346 laAspProProAlaAlaAlaArgLeuProGluArgGlnGluProArg 362
947 ...CCTCGGTGTGTAGCCCCCGC...CACTGTCCACTGCGCTGCGCCC 904
      |||:|:| ||| :|:|:|
363 LeuProGlnAspLeuAlaAlaGlnArgCysProAlaGly.....Pr 377
903 TCAGAGCCCCAGCTGGGGGAGCGGCTGC.....AGGAGCGCT 866

```



```

13 LeuGlnSerThrAlaIleLeuSerLeu.....Va 22
319 AGCTGGGATGCTACTACTAGG.....G 341
|||||.....|
22 IAlaAlaGluAlaThrGlnThrGlnTyrGlyGlyTyrLeuProProValT 39
342 AGGTCTGCTGGGCGCTGAGGTGTGCAGCTCAGAGCCCGCTGAGGCC 391
|||||.....|
39 hrSerGlnProProSerSerIleGlyLeuSerProProSerAlaPro 55
392 CCAGGAACCCCTCTCCCTCCACAGCCATCTAGATGAGCTCCCTCTGCC 441
|||||.....|
56 ThrThrThrProPro...SerArgGlyHisVal.....ProSerPr 68
442 CACTGTGTCTACTCTAAAGAGCTTCAGAGAGGCTCTCCAGAGCGGCT 491
|.....|
68 o.....ArgHisAlaProProArgHisAlaT 77
492 GGGACTCGGACCAATGCACTCCAAGCGTCTTTGGGGACCCCTGCC 541
::.....|
77 yrProProProSerHisGlyHisLeuProProSerValGlyGlyPro... 92
542 ACTCCACGGAAGGT.....CCTCCCGCG.....CCACCTCC 573
|||.....|
93 ProProHisArgGlyHisLeuProProSerArgGlyPheAsnProProPr 109
574 CGTCTCTCC.CTGTGCGGTTCAGATGGCGCGCGTCACAACTCTGTAGC 622
|.....|
109 oSerProValIleSerProSerHisProProSerTyrGlyAla...P 125
623 CTGTGACCC.....GGCCGAGGCCAA...CAACAGTGAAGGCGGACCA 663
||.....|
125 roProProSerHisGlyGlyGlyHisLeuProProSerHisGlyGlnArgPro 141
664 TGACTACCTGCGCTGCTGGCGGT.....GCAGGAGGAC..... 698
|||||.....|
142 ProSerProSerHisGlyHisAlaProProSerGlyGlyHisThrProPr 158
699 .....CAGGTCCTTCGCGTGGACGCGCGCTCTTGGCGCGCTGTCG 741
|||.....|
158 oArgGlyGlnHisProProSerHisArgProProPro..... 172
742 CATCTGCGAGGAC.....GCTCTGCGGTGCT 770
|||||.....|
173 ..SerArgHisGlyHisProProProThrThrAlaGlnProProPro 188
771 CGCCCTTTCGCTGACCGCGCGCGCTCAGCCCGCTCGGCGCTCATG 820
|||||.....|
189 ThrProIleTyrSerProSerProGlnValGlnProProProThrTyrSe 205
821 CCTGGGCACTTCTGGGCGCTGCTGTTGGCCAGGTTGACAGTGCACGCG 870
|||.....|
205 rPro.....ProProP 209
871 CTCTGCGAGCCGCTCCCGAGCTCGGGCTCTGAGGGCGCGGAGGTGG 920
||.....|
209 roThrHisValGlnPro..... 214
921 ACAGTGGGCGGCGCTCAGACACCGAGGCTCCGATGGGCGGAAGGCTG 970
|||||.....|
215 .....ThrProSerProProSer.ArgGly..... 222
971 GCGGCGCAGCACTTGGGCGCGGAGCTGGGCGACTGCGCTGACACTCGC 1020
|||||.....|
223 .....HisGlnProGlnProThrHisArgHisAlaProProThrHis 237
1021 CTGGCTGGAGCAGGATGCGCGCTGCTTCGACGAGTGGGACGTGACAG 1070
::.....|
238 ArgHisAlaProProThrHis..... 244
1071 CGGCCAAGGCTGATTCTGCTGCGGCGCGGCGGCGGCGCTTGCCTGACGCGCT 1120
|||.....|
245 .....GlnProS 247

```

```

1121 GACCTGGCGCGCTCAAGGCGCAGCGGAGGCTCTTCTGCTACTGCG 1170
|||.....|
247 erProLeuArgHisLeuProProSerProArgGlnProGlnProPro 263
1171 CCACTGGGACCAAACTGCGAGGTACACCTGCTGTGTACAGCCCGCAGGA 1220
|||||.....|
264 ThrTyrSerProProProAlaTyrAlaGlnSerProGlnProSerPr 280
1221 ACGTCTGAAGGCTCCCGCTGCTTGGCTGGCGCGCGCCACCAACACA 1270
::.....|
280 oThrTyrSerPro.ProProProThrTyrSerProProProSerPro 296
1271 CTCAGTCACTGCGCGCGAGGCTGCTTGTGTGCTGGGAAAGTGTAG 1320
::|||.....|
297 IleTyrSerProProProAla..... 304
1321 GCTGCTGCGACGCTGCTCCCGCTCTTCTTACTCTCCCTGCGGAGCGCT 1370
|||||.....|
305 ....TyrSerProSerProProProThrProThrProThrPheSerProp 320
1371 CTGCGCGCGCACAAAAGTGCCTGCTGTCTCTCCCTCTCTCTCCCGCACC 1420
|||.....|
320 roProProAlaTyrSerProProProThrTyrSerProProProProThr 336
1421 CCACTCACTCCCGCTCC 1438
|||.....|
337 TyrLeuProLeuProSer 342
seq_name: SwissProt_40:NME4_MOUSE
seq_documentation_block:
ID NME4_MOUSE STANDARD; PRT; 1323 AA.
AC Q03391;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
GN GRIN2D.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93050214; PubMed=1365220;
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RI Inoue Y., Mishina M.;
RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor
channel."
RL FEBS Lett. 313:34-38(1992).
RN [2]
RP REVISIONS.
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA Inoue Y., Mishina M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC -1- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```



```

1169 .....||| |
1225 GFGAAGGCTGCCCTGCTTGGCTGGCGCCCAACACACACTCA 1274
1171 yALaSerLeuGlulLeu.....ProProArgHIsLeu 1183
1275 AGTCACTGCCGCCAGGCTGGCTCTTGGTGTGGAAAGTGTAGGCTG 1324
1184 SerCysSerHisAspGlyLeuAspGly.....GlyTr 1194
1325 GTGCCAGCTGTCCCCCACTGCTTCTTACTCCCTCCCTGAGCCCTCTTG 1374
1194 pTrpAlaProPro.....ProProTrpAlaAlaGly 1206
1375 CCCCCACAAAAGTGCCTGCCTGTGCTCTCTCTCTCTCTCTCTCTCCACCCAC 1424
1206 roProProArgArgAlaAlaGlyCysPro.....ArgProHis 1220
1425 TCACACTCCCTCCATCTCT 1445
1221 ProHisArgProArgAlaSer 1227
seq_name: SwissProt_40:EBN1_EBV
seq_documentation_block:
ID EBN1_EBV STANDARD; PRT; 641 AA.
AC P03211;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-DEC-1998 (Rel. 37, Last annotation update)
DE EBNA-1 nuclear protein.
DE BKRF1.
GN Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_taxid=10377;
[1]
RP SEQUENCE FROM N.A.
RP MEDLINE=84270667; PubMed=6087149;
RX Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell
RX Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Sequi
RX Tuffnell P.S., Barrall B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus
RL Nature 310:207-211(1984).
[2]
RP SEQUENCE OF 1-26 FROM N.A.
RP MEDLINE=86259739; PubMed=3460083;
RX Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RA "Nucleotide sequences of mRNAs encoding Epstein-Barr virus
RT proteins; a probable transcriptional initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
[3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90266473; PubMed=2161150;
RX Petti L., Sample C., Kieff E.;
RA "Subnuclear localization and phosphorylation of Epstein-Bar-
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
[4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX MEDLINE=96006523; PubMed=7553871;
RX Bochkarev A., Barwell J.A., Puettner R.A., Furey W.F. Jr.,
RA Edwards A.M., Frappier L.;
RT "Crystal structure of the DNA-binding domain of the Epstein
RT virus origin-binding protein EBNA 1.";
RL Cell 83:39-46(1995).
CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN
CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSACTIVATING
CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM,
CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASS
CC WITH THE NUCLEAR MATRIX.

```


537 ..SerArgLeuPro 540
|||||

seq_name: SwissProt_40:NME4_HUMAN

seq_documentation_block:

ID NME4_HUMAN STANDARD; PRT; 1336 AA.
AC O15399;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
GN D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D) (EB11).
GN GRIN2D.
OS Homo sapiens (Human).
ID Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98149394; PubMed=9489750;
RA Hess S.D., Daggett L.P., Deal C., Lu C.-C., Johnson E.C.,
RA Velicelebi G.;
RT "Functional characterization of human N-methyl-D-aspartate subtype
1A/2D receptors.";
RL J. Neurochem. 70:1269-1279(1998).
CC -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC WITH HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC SENSITIVITY TO MAGNESIUM. MEDIATED BY GLYCINE.
CC -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
CC send an email to license@isb-sib.ch).
CC
CC EMBL; U7783; AAC15910.1; .
CC HSP; P19491; 1GR2.
CC MIM; 602717; .
CC InterPro; IPR001320; Ion_glut_receptor.
CC InterPro; IPR001311; SBP_glut_receptor.
CC Pfam; PF00060; lig_chan; 1.
CC SMART; SM00079; PBPe; 1.
CC Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW Glycoprotein; Ionic channel; Magnesium.
FT SIGNAL 1 27
FT CHAIN 28 1336
FT FT
FT FT
FT DOMAIN 28 583
FT TRANSMEM 584 604
FT FT
FT DOMAIN 605 622
FT TRANSMEM 623 641
FT DOMAIN 642 657
FT TRANSMEM 658 682
FT DOMAIN 683 846
FT TRANSMEM 847 865
FT DOMAIN 866 1336
FT DOMAIN 281 286
FT DOMAIN 908 916
FT DOMAIN 1035 1040
FT DOMAIN 1209 1213
FT DOMAIN 1244 1247
FT SITE 642 642
FT FT
FT CARBOHYD 92 92
FT CARBOHYD 352 352
FT CARBOHYD 366 366

FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 642 642 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1336 AA; 143558 MW; DECC545F3E416680 CRC64;

alignment_scores:
Quality: 200.00 Length: 455
Ratio: 1.258 Gaps: 24
Percent Similarity: 34.945 Percent identity: 24.615

alignment_block:

US-09-786-136-4 x NME4_HUMAN

Align seg 1/1 to: NME4_HUMAN from: 1 to: 1336

353 GGGGGCCCTGCAGGTGTGCAGCTCAGAGCCGCTGAGCCGCCAGCCAGCAACCC 402
|||
892 GlyMetTyrSerCysSerAlaGluAlaProProProAlaLysPr 908
403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCTCTGCCACTGTGTGTCT 452
|||
908 oProPro.....ProProGlnProLeuProSer..... 917
453 ACTCTAAGGACTTTCAGAGAGCTCTCCAGCAGGCGCTGGGACTCGGAC 502
917 917
503 CAAATGGCACTCCCAAGCGTCTTTGGGGACCTGCCACTCCACCGGA 552
|||
918ProAlaTyrProAlaPr 923
553 AGGTCTCGCCGCCACCT.....CCCCCTCTCCCTGTCTGCGC 590
|||
923 oGlyProAlaProGlyProAlaProPheValProArgGluArgAlaSerV 940
591 TCAGCATGGCGCGCGCTCACAACTCTGTAGCCCTGACCCGGCGGCGGCC 640
|||
940 AlaArgTrpArgargThrLysGlyAlaGlyPro...ProGlyGlyAla 955
641 AACACAGTGAAGCAGGACGACCTACTGCTGCTGCTGCTGCTGCTGCTG 679
|||
956 GlyLeuAlaAspGlyPheHisArgTyrTyrGlyProIleGluProGlnGl 972
680GTGCGGCTGCAGGAGGACCAAGCTCTTCGCTGCGACGCGC 722
|||
972 yLeuGlyLeuGlyLeuGlyGluAla.ArgAla..... 982
723 CCTTCTGCGCGCTGTGCGCATCTCGCAGGAGCGGCTCTGCGCTGCTCG 772
|||
983AlaProArgGlyAlaAlaGlyArg..... 990
773 CCTTTGCGTGCACCGCGCGCAGCTCAGCCGCCAGCTCGGCTCATTCGC 822
|||
991ProLeuSerProProAlaAlaGlnProProGlnLysPro...P 1004
823 CTGGGCACTCTGCGCGCTGTGTTGGCCAGGGTGACAGTG..... 863
|||
1004 roAlaSerTyrPheAlaIleValArgAspLysGluProAlaGluProPro 1020
864CCAGCGCTCTCGCAGGCGCGCTGCTGCTGCTGCTGCTGCTG 892
1021 AlaGlyAlaPheProGlyPheProSerProProAla.....ProProAl 1035
893 TCGGCTCTGAGG..... 905
1035 aaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052
906GGCCAGGCGGAGGTGACA..... 923
1052 spGluSerProProAlaProAlaArgTyrProArgSerAspProGluSer 1068
924GTGGCGGGGCTCAGACACCGGCGCTCC 952

```

1069 GlnProLeuLeuGlyProGlyAlaGlyGlyAlaGlyGlyThrGlyGlyAl 1085
953 GATGGGGGGAAGGCTGGCGCCACACCTGCGGGCCGAGCCTGGCC 1002
1085 aGlyGlyGlyAlaProAlaAlaProProCysCysAlaAlaProProp 1102
1003 CACTGCCG..... 1010
1102 roCysProTyrLeuAspLeuGluProSerProSerAspSerGluAspSer 1118
1011 .....TAGCACTCGCTGCTGGAGCAC 1033
1119 GluSerLeuGlyGlyAlaSerLeuGlyGlyLeuAspProTyrPheAl 1135
1034 CGATCGCGCGCTGCTTCGACGAGTGGGAAGTACGACGGCCAGGCTGA 1083
1135 aAspPheProTyrProTyrAlaGlu.....ArgLeu. 1145
1084 TTGCTGGCTGGCGGCCGACGACTTGCCTGACGGCTTGACCTGGCGGCC 1133
146 .....GlyPro.....ProPro 1149
1134 TCAAGS.....CCGACGCCGAGGCTCTTC 1159
1150 GlyArgTyrTrpSerValAspLysLeuGlyGlyTrpArgAlaGlySerTr 1166
1160 CTGCTACTCGCGCACTGGGACCAAACTGCGAGCTACACCTGC.....T 1203
1166 pAspTyrLeuProProArgSerGlyProAlaAlaTrpHisCysArgHisC 1183
1204 GTGCTACAGCCCGCAGGAGCTGTGAAGCTGCGCCCTGCTGCTGGGCTG 1253
1183 ysAlaSer.....LeuGluLeu 1188
1254 GCGCCCCCAACACACTCAAGTCACTGCGCCGCGGCTGGCTCTTG 1303
1189 LeuProProProArgHisLeuSerCysSerHisAspGlyLeuAspGly.. 1204
1304 GTGCTGGGAAAGTGTAGCTGTGCGCAGCTGTCCGCCACCTGCTTCTTAC 1353
1205 .....GlyTriPAlaProProPro..... 1211
1354 TCCTCTCCCTGGAGC.....CCTCTTGGCCCAACAAAGTGGCTGCTG 1397
1212 ..ProProTriPAlaAlaGlyProLeuProArgArgAlaArgCysGly 1227
1398 TGC.....TCCTCTCCCTGCTCCGCCACCTCACACTCCCTCCATC 1441
1228 CysProArgSerHisProHisArgProArgAlaSerHisArgThrProAl 1244
1442 CTCTGAGCTCCCT 1454
244 aAlaAlaPro 1248
Seq_name: SwissProt_40:TEGU_EBV

```

```

seq_documentation_block:
ID TEGU_EBV STANDARD; PRT; 3149 AA.
AC P03186;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein.
GN BPLF1.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,

```

```

RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Sequin C.,
RT Tuffnell P.S., Barrell B.G.;
RL "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC 1- FUNCTION: TEGUMENT PROTEIN
CC 1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHVS-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01555; CAA24839.1; -.
CC PIR; A03747; Q08E8.
CC PIR; S32993; S32993.
CC SEQUENCE 3149 AA; 337954 MW; 3DD0C576587313D8 CRC64;

```

alignment_scores:

```

Quality: 199.50 Length: 415
Ratio: 1.108 Gaps: 23
Percent Similarity: 43.373 Percent Identity: 26.988

```

alignment_block:

```
US-09-786-136-4 x TEGU_EBV ..
```

```
Align seg 1/1 to: TEGU_EBV from: 1 to: 3149
```

```

483 CAGCGCGCTGGGACTCGGACCAAACTCCAAAGCGTCTTTGGG 532
279 GluAlaAlaGlyThrProGlyAlaAspSerProValMet..... 293
533 GACCTGCCACTCCACGGAAGTCTTCGCGCCGCGCCCTCCCGCTCTCC 582
294 ...LeuProPheGluArgArgilleleProTyrAsnLeuArgProLeuP 309
583 CT.....GTGCGCTCAGCATGGCGCGCTCACAACTCTGTA 620
309 roSerArgSerPheThrSerAspSerPheProAlaAlaArgTyrSerPro 325
621 GCCCTGACCGCGCCAGGCCAACAA...GTGAAGCAGCAGCATGAC 667
326 AlalysThrAsnSerProProSerSerProAlaSerAlaAlaProAlaSe 342
668 TACCTGCCCTTGGTGGGCTGAGGAGGACAGGCTCTTCGCGCTGGA 717
342 rAlaAlaProAlaSerAlaAlaProAlaSerAlaAlaProAlaSer.... 357
718 CGCGCCCTTGTGGCGCTGTGGCGCATCTCGCAGGAGCGCTCTGCGGTG 767
358 .....AlaAlaPro.....AlaSerAlaAlaProAlaSerAlaAla 369
768 CCGCGCCCTTGTGGCGTGCAGCGCCGCGCTCGACCCACCT..... 809
370 Pro.....AlaSerAlaAlaProAlaSerSerProProLeuPheII 383
810 .....CGGCTCATTTGCCCTGGGCTGCTTGGGCGCTGTGT 846
383 eProileProGlyLeuGlyHisThrProGly..... 393
847 TGGCCAGGCTGACAGTGCACGCGCTCTCGACCGCTCGCCGCTCCCGCAGCTCGG 896
394 .....ValProAlaProSerThrProProArg 402
897 GCTGTAGGGGCCAGGCCAGTGTGGCGGGCTCAGACACCGAG 946
403 AlaserSerGlyAlaAla.....ProGI 410
947 GCCTCCCATGGGCGGAGGCTGGGCGG.....GCAC 978

```

```

:      |||      ::|||::|      |
410 nProLysArgLysLysGlyLeuGlyLysAspSerProHisLysLysP 427
979 CGACCTCGGGGGGACCTGGCCACTGCGTAGCAGTCGCGCTGCTGG 1028
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
427 roThrSerGlyArg.....ArgLeuProLeuSerSerThrThrAsp 440
1029 AGCAGCGATCGCGCTGCTTCCAGAGCTGGGAAGTACAGCGGCCAAG 1078
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
441 ThrGluAspAspGlnLeuProArgThrHisValProProHisArgPro 457
1079 GCTGATTGCTGGTGGGGGCCAGCACTTGCCTGAGCGGCTTACCTGGC 1128
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
457 oSerAlaAlaArgLeuProProValIleProIleProHisGlnSerP 474
1129 CGGCTCAAGCGCGAGCGCGAGG.....1152
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
474 roProAlaSerPro.ThrProHisProAlaProValSerThrIleAlaPr 490
1153 .....1159
oSerValThrProSerProArgLeuProLeuGlnIleProIleProLeuP 507
1160 CTGCTACTGCGCCAGCTGGGACCAAAA...CCTGCAGCTACACCTGCTGTG 1206
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
507 roGlnAlaAlaProSerAsnProLysIleProLeuThrThrProSer... 522
1207 CTACAGCCAGCAAGCTGTGAGGCTGCGCCCTGCTGCTGTGGCTGGCG 1256
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
523 .....ProSerProThrAlaAlaAlaAla.ProThrThrThrThrLeuS 537
1257 CCCACCCACACACACCACTCACTGCGCCAGGCTGGCTGCTGTGGTG 1306
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
537 erProProProThr...GlnGlnProProGlnSerAla..... 549
1307 CTGGGAAGTGTAGCTGTGGCGGCTGTGCCCTGCTGCTT..... 1348
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
550 .....AlaProAlaProSerProLeuLeuProGlnGln 560
1349 ....CTTACTCCCTCCCTGGAGCCCTCT..... 1372
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
560 nGlnProThrProSerAlaAlaProAlaProSerProLeuLeuProGlnG 577
1373 .....TGCCTCCAC 1381
577 lnGlnProProProSerAlaAlaArgAlaProSerProLeuProProGln 593
1382 AAAAGTGCCTGCTGTGCTCTCTCCCTCCCTCCACCCACCTACACAT 1431
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
594 GlnGlnProLeuProSerAlaThrProAlaPro.ProProAlaGlnGlnL 610
1432 CCCCTCATCTCTGAGCTC.....CCTGCA 1457
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
610 euProProSerAlaThrThrLeuGluProGluLysAsnHisProProAla 626
1458 ACACAGTGAAGGTAGAGAGCCACAGTCCCAATCCT 1496
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
627 AlaAspArgAlaGlyThrGlu...IleSerProSerPro 638

```

seq_name: SwissProt_40_YHL1_EBV

seq_documentation_block:

```

ID YHL1_EBV STANDARD; PRT; 660 AA.
AC F03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLFI protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Gammaherpesvirinae; no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]

```

```

SEQUENCE FROM N.A.
MEDLINE=84270667; PubMed=6087149;
BAER R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
Tuffnell P.S., Barrell B.G.;
"DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01555; -; NOT_ANNOTATED_CDS.
CC DR PIR; A03742; Q0BE3.
CC KW Hypothetical protein; Early protein; Repeat.
CC FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
CC FT REPEAT 149 273 1.
CC FT REPEAT 274 398 2.
CC FT REPEAT 399 523 3.
CC FT REPEAT 524 648 4.
CC SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

```

alignment_scores:

```

Quality: 196.50 Length: 480
Ratio: 0.949 Gaps: 30
Percent Similarity: 43.125 Percent Identity: 27.708

```

alignment_block:

US-09-786-136-4 x YHL1_EBV ..

Align seg 1/1 to: YHL1_EBV from: 1 to: 660

```

350 CTGGGGGCTCAGGTGGAGC.....TCAGAGCCGCTGA 387
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|
172 ProGlyAlaGlyGlnArgProSerGlyProThrGlyArgProAlaAl 188
388 GCCC.....CCAGAACCCCTCCTCCCTCCACAGCCATCATGTCAG 431
|||::|::|::|::|::|::|::|::|::|::|::|::|::|
188 aProGlyAlaProGlyThrPro.....AlaAla 197
432 CTCTCTGCCACTGTGTCTACTTAAGGACTTCAGAGAGGCTCTCCA 481
|||::|::|::|::|::|::|::|::|::|::|::|::|::|
197 laProGlyPro.....GlyGlyAlaAlaValPro 207
482 GCAGGGCCTGGGACTCGGACCAAAATGCCAACTCCAAGCTGCTTGGG 531
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
208 SerGlyAlaThrProHisProGluArgGlySerGly..... 219
532 GGACCTCGCCACTCCACCGAAGGTCCTCGCCGCCACCT..... 571
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|
220 ....ProAlaAspProProAlaAlaAlaArgLeuProGluArgGlnG 235
572 ..CCCGCTCCTCCTGCTGGCTCAGCATGGCGCCGTCACAACTCTGT 619
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|
235 luProArgLeuProGlnAspLeuAlaAlaGlnArg.....Cys 248
620 .....AGCCCTGACCCGGCCAGGCAACACAGTGAAGGAGGAGCA 663
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
249 ProAlaGlyProProPro.ThrArgSerGlyAlaAlaGlnArgThrH 265
664 TGACTACCTGCTTGGTGC.....GGCTGC... 689
265 isArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysPro 281
690 .....AGGAGGACACGAGGCTCCTTC 709
|||::|::|::|::|::|::|::|::|::|::|::|::|::|
282 ArgThrThrArgArgSerGlyAlaGlnArgGlyHisProProGln 298

```


OM of: US-09-786-136-4 to: Issued_Patents_AA: * out_format : pfs
Date: Sep 19, 2002 5:42 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODE=framer_n2p.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09786136/runat_16092002_141242_19186/app_query.fasta_1.1592
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-LOOPEXT=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-NS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pcp
-MAX=100 -THR_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs
-M=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USERS=US09786136 @CGL_1.14 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-786-136-4
Query length: 1511
Database: Issued_Patents_AA: *
Database sequences: 2316298
Search time (sec): 34.210000

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-026-138E-4	+	205.50	248.07	1.8e-07	1323
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-483-533-41	+	193.50	245.10	1.0e-06	355
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-283-471A-41	+	193.50	245.10	1.0e-06	355
/cgn2_6/ptodata/1/iaa/PTCUS_COMB.pcp:US91-06532-3	+	193.50	245.10	1.0e-06	355
/cgn2_6/ptodata/1/iaa/5B_COMB.pcp:US-08-231-193A-58	+	191.00	228.93	2.1e-06	1336
/cgn2_6/ptodata/1/iaa/5B_COMB.pcp:US-08-486-273A-58	+	191.00	228.93	2.1e-06	1336
/cgn2_6/ptodata/1/iaa/6A_COMB.pcp:US-08-940-086A-58	+	191.00	228.93	2.1e-06	1336
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-940-035A-58	+	191.00	228.93	2.1e-06	1336
/cgn2_6/ptodata/1/iaa/5B_COMB.pcp:US-08-316-650-12	+	189.50	226.22	2.8e-06	1442
/cgn2_6/ptodata/1/iaa/PTCUS_COMB.pcp:US95-02251-12	+	189.50	226.22	2.8e-06	1442
/cgn2_6/ptodata/1/iaa/6A_COMB.pcp:US-08-906-865-4	+	189.00	232.64	2.5e-06	696
/cgn2_6/ptodata/1/iaa/6A_COMB.pcp:US-08-931-820-3	+	188.50	227.90	3.1e-06	1060
/cgn2_6/ptodata/1/iaa/6A_COMB.pcp:US-08-963-825-2	+	188.50	225.07	3.3e-06	1418
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-010-993-1	+	188.50	225.07	3.3e-06	1418
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-570-573-20	+	188.50	225.07	3.3e-06	1418
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-548-608-20	+	188.50	225.07	3.3e-06	1418
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-07-945-288-15	+	187.00	219.96	4.6e-06	1958
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-313-289B-15	+	182.00	227.29	7.5e-06	469
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-209-747-2	+	179.50	218.43	1.3e-05	832
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-458-298-2	+	179.50	218.43	1.3e-05	832
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-103-429A-4	+	178.50	217.44	1.5e-05	805
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-556-978B-19	+	177.00	217.53	1.9e-05	651
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-247-806-1	+	177.00	217.53	1.9e-05	651
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-425-063-2	+	177.00	216.58	1.9e-05	718
/cgn2_6/ptodata/1/iaa/5B_COMB.pcp:US-08-317-844B-2	+	177.00	216.58	1.9e-05	718
/cgn2_6/ptodata/1/iaa/6A_COMB.pcp:US-09-034-177-3	+	177.00	216.19	2.0e-05	747
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-219-849-4	+	175.50	213.29	2.6e-05	822
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-628-823-4	+	174.50	205.54	3.6e-05	1593
/cgn2_6/ptodata/1/iaa/5B_COMB.pcp:US-08-660-963-12	+	174.00	216.19	2.9e-05	498
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-219-849-5	+	173.50	209.15	3.7e-05	960
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-556-978B-63	+	173.00	213.00	3.6e-05	604
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-556-978B-21	+	172.00	211.66	4.3e-05	606
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-247-806-4	+	172.00	211.66	4.3e-05	606
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-09-219-849-4	+	172.00	209.98	4.5e-05	720
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-642-255-53	+	172.00	209.24	4.6e-05	777
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-046-583-5	+	172.00	199.88	5.8e-05	2035
/cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-393-703-5	+	172.00	199.88	5.8e-05	2035
/cgn2_6/ptodata/1/iaa/PTCUS_COMB.pcp:US93-11721-5	+	172.00	199.88	5.8e-05	2035
/cgn2_6/ptodata/1/iaa/6B_COMB.pcp:US-08-556-978B-59	+	171.00	212.35	4.8e-05	493
/cgn2_6/ptodata/1/iaa/6A_COMB.pcp:US-08-963-825-19	+	171.00	202.44	6.2e-05	1366

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pcp:US-08-026-138E-4

seq_documentation_block:
; Sequence 4, Application US/08026138E
; Patent No. 5502166
; GENERAL INFORMATION:
; APPLICANT: Masayoshi MISHINA
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nishiohata Residence 1-107
; STREET: 5214, Nishiohata-machi
; CITY: Niigata-shi
; STATE: Niigata-ken
; COUNTRY: JAPAN
; ZIP: 951
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS v.5
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/026,138E
; FILING DATE: 26-FEB-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 39563/1992
; FILING DATE: 26-FEB-1992
; APPLICATION NUMBER: JP 173155/1992
; FILING DATE: 30-JUN-1992
; APPLICATION NUMBER: JP 215017/1992
; FILING DATE: 12-AUG-1992
; APPLICATION NUMBER: JP 303878/1992
; FILING DATE: 13-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hamburg, C. Bruce
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-4551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 986-2340
; TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1323 amino acids
; TYPE: amino acid
; STRANDEDNESS: single strand
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: mouse
; TISSUE TYPE: brain
; PUBLICATION INFORMATION:
; AUTHORS: Masayoshi MISHINA
; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
; RELEVANT RESIDUES IN SEQ ID NO: 4: FROM 1 TO 1323
; US-08-026-138E-4

alignment_scores:
Quality: 205.50 Length: 443
Ratio: 1.309 Gaps: 26
Percent Similarity: 35.440 Percent Identity: 26.185

alignment_block:
US-09-786-136-4 x US-08-026-138E-4
Align seg 1/1 to: US-08-026-138E-4 from: 1 to: 1323

TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-533-41

alignment_scores:
Quality: 193.50 Length: 378
Ratio: 1.166 Gaps: 22
Percent Similarity: 43.915 Percent Identity: 28.571

alignment_block:
US-09-786-136-4 x US-08-483-533-41 ..

Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355

```

353 GGGGCGCTCGAGGTGTCAGCTCAGAGCCGCTGAGCCCGCCAGGAAACCC 402
|||||: : : : : : : : : : : : : : : : : : : : : : : : :
20 GlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSerG1 36
403 TCGTGGCTTCACAGCACTATGATGCGAGCTCCTCTGCCC..... 442
|||||: : : : : : : : : : : : : : : : : : : : : : : :
36 uProAlaValArgSerProAlaAlaAlaProProProProAlaG 53
443 .....ACTGTGCTACTCTAAAGGACTTCAGAGA 472
53 lyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis... 68
473 GGCTCTCCAGCAGGCGCTGGGACTCGGACCAAAATGCAACTCCAAGCG 522
|||||: : : : : : : : : : : : : : : : : : : : : : : :
69 ...ValProGluSerAlaSerAspAspAspAspAspAspAspAsp 84
523 TGTCTTTGGGGACCCCTCCACTCCCGAGGAAGTCTCGCCGCCCCACCTC 572
|||: : : : : : : : : : : : : : : : : : : : : : : :
84 pSerProProProGluSerAlaProGluAlaArgProThrAlaAlaAlaP 101
573 CCGTCTCTCCCTCGCTCGGCTCAGCTGCGCGCGGCTCACAACTCTGTAGC 622
|||||: : : : : : : : : : : : : : : : : : : : : : : :
101 roArgPro..... 103
623 CTGACCCGGGCGAGCCCAACACAGTGAAGCAGCGACCATGACT.... 668
||| ||| ||||||: : : : : : : : : : : : : : : : : :
104 ProGlyPro.HisArgProAlaTrpAlaArgGlyAlaGlyLeuThrProp 120
669 ..ACCTGCCCTTGGTGGCTCGAGGAGCAGGCTCTTCCGCTGG 716
||| ||| : : : : : : : : : : : : : : : : : : : : : :
120 roThrProPro.....ArgAlaProSerAlaPhe 129
717 ACAGCGCTCTTCCGCGCTGCGCATCTCCAGGAGCGCTCTGCGCT 766
||| ||| ||||||: : : : : : : : : : : : : : : : : :
130 ArgArgAlaSerProSerAlaCysAlaSerProArgSerThrTrp.... 144
767 GCCTCGCCCTTGGCTGCGCGCGGCGGCGGCTCA.....GCC 804
||| ||| ||| ||| |||: : : : : : : : : : : : : : :
145 ....ArgAlaCysAlaCysAspAlaArgAlaGlyArgGlyArgSerP 160
805 CACCTCGGCTCATGCCCTGGGCACTTCTGGGCCCTGTGTGGCCAGG 854
|||||: : : : : : : : : : : : : : : : : : : : : :
160 roProArgProProArgPro..... 166
855 GTGACAGTGCAGGCGCTCT...GCAGCGCGCTCCCGCA..... 890
||| ||| ||||||: : : : : : : : : : : : : : : : :
167 .....ProArgProProArgProProArgProProArgGlyCysAl 180
891 .GCTCGGCTCTGAGGGGCGAGCCAGGTGGACAGTGGGCGGCTCAGA 939
||| ||| ||| ||| |||

```

```

180 aSerArgProThrSerGlyCysAlaThrTrpTrpSerGly..... 193
940 CACCGAGCCTCCGATGGGCGGGAAGGCTGGGCGGCGACGACCTGCGGG 989
|||||: : : : : : : : : : : : : : : : : : : : : :
194 ..ProArgProProAlaTrpArg..AlaAlaAlaArgGlyProAlaSerG1 209
990 GCGG...GACCTGGGCGCACTGCCGTAGCAGCTCGCTGGTGGAGCAGGA 1036
||||| ||||||: : : : : : : : : : : : : : : : :
209 yProThrGlyLeuGly.....SerGlyAlaGlyT 219
1037 TGGCGCGCTGCTTCGACGAGTGGAACTGACAGCGGCGCAAGGCTGATTG 1086
||||| ||| |||: : : : : : : : : : : : : : :
219 rpArgArgProArgArgSerSerGlyArg..... 228
1087 CTGGCTGGGCGGCGGAGCTGCTGAGCGGCTTGCCTGACCTGGCGGCTCA 1136
||| ||||||: : : : : : : : : : : : : : : : :
229 ...AlaTrpGlyProArgProValProGlyProTrpProAlaGluProAl 244
1137 AGGCGCG.....AGCCGAGGCT.....CTTCTCTGCTA 1165
||| ||| |||: : : : : : : : : : : : : : :
244 aArgArgThrArgSerAsnValThrProGluAlaAlaTrpValPheArg 261
1166 CTGGCGCACTGGGACCAAACTGCAGCTACACCTGCTGTCTACAGCCC 1215
|||||: : : : : : : : : : : : : : : : :
261 lyAlaProGlySerSerAlaProSerArgSerProGluArgArgTrpGln 277
1216 AGCGAAGCTGGAAGGCTGCCCTGCTGCTGGGCTGGGCGGCGGCGGCA 1265
||| ||||||: : : : : : : : : : : : : : : :
278 GluProArgGlyThr.....LeuGlyAlaSerProProse 290
1266 ACACACTCAAGTCACTGCCGCGGCGGCTGCTTGGTGGTGGGAAG 1315
||| ||| |||: : : : : : : : : : : : : : :
290 rGlnGlyGlyProProArgGlyArgAspTrpProIleGly...GlyArg 306
1316 TGTAGGCTGGTCCAGCTGCTGCCCTGCTTCTTACTCCTCCTCCTGGA 1365
||| ||| |||: : : : : : : : : : : : : : :
306 lnArgGlyGlyAla.....ArgProThrSerValArgValPheGly 320
1366 GCGCTC.....TGGCGGCGCA 1380
||| ||| |||: : : : : : : : : : : : : :
321 AlaArgProIleGlyArgGluLeuProPro 330

```

```

seq_name: /cgn2_6/ptodata/1/1aa/68_COMB.pep:US-09-283-471A-41
seq_documentation_block:
; Sequence 41, Application US/09283471A
; Patent No. 6340673
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/283,471A
; FILING DATE: 04-APR-1999
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853

```



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.193A
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,459
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; 8-231-193A-58

```

```

alignment_scores:
  Quality: 191.00      Length: 450
  Ratio: 1.194        Gaps: 24
  Percent Similarity: 35.556  Percent Identity: 24.667

```

```

alignment_block:
US-09-786-136-4 x US-08-231-193A-58 ..
Align seg 1/1 to: US-08-231-193A-58 from: 1 to: 1336

```

```

353 GGGGCCCTGAGGTGTGACGCTGAGCCCGCTGAGCCCGCCAGGACCC 402
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
892 GlyMetTyrSerCysCysSerAlaGluAlaAlaProProAlaLysPr 908
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
403 TCCTGCCTCTCACAGCATCATGATGAGCTCTCTGCCACACTGTGTCT 452
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
908 OPTRO.....ProProGlnProLeuProSer..... 917
453 ACTCTAAGGACTTCAGAGAGGCTCTCCAGCAGGCCCTGGGACTCGGAC 502
917 ..... 917
503 CAANAATGGCAACTCCAAGCGTGTGTTGGGGAGCCCTGCCACTCCACGGA 552
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
918 .....ProAlaTyrProAlaPr 923
53 AGTTCCTCGCGCCGACCT.....CCCGTCTCTCCCTGTCGCGC 590
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
23 oGlyProAlaProGlyProAlaProPheValProArgGluArgAlaSerV 940
591 TCAGCATGGCCCGCGTACAACTCTGTAGCCCTGACCGCGCCAGGCC 640
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
940 alAlaArgTTPArgProLysGlyAlaGlyPro...ProGlyGlyAla 955
641 AACAAAGCTGAAGCAGCGACCATGACTACTGCTGCTG..... 679
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
956 GlyLeuAlaAspGlyPheHisArgTyrTyrGlyProIleGluProGlnG 972
680 .....GTGCGGCTGAGGAGGACGAGCTCTCTCCGCTGAGCGCGC 722
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
972 yLeuGlyLeuGlyGluAla.ArgAla..... 982
723 CCTTCTCGCGCGTGTGCGCATCTCCAGGAGCGCCCTGCTGCCGCTGCTCG 772
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
983 .....AlaProArgGlyAlaAlaGlyArg..... 990
773 CCCTTTCCGCTGACCGCGCGCCGCTCAGCCCGCCCTGCGCTCATGTGCC 822

```

```

  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
991 .....ProLeuSerProAlaAlaGlnProGlnLysPro...P 1004
823 CTGGGCACTTCTGGGCGCTGGTGGCCAGGCGTGCACACTG..... 863
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1004 roAlaSerTyrPheAlaIleValArgAspLysGluProAlaGluProPro 1020
864 .....CCAGGGCTCTGCGAGCGCGCTCCGCCAGC 892
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1021 AlaGlyAlaPheProGlyPheProSerProProAla.....ProProAl 1035
893 TCGGCTCTGAGG..... 905
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1035 aAlaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052
906 .....GCCAGCGCCAGGTGCACA..... 923
1052 spGluSerProProAlaProAlaArgTTPProArgSerAspProGluSer 1068
924 .....GTGGCGGGGCTCAGACACCGAGGCTCC 952
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1069 GlnProLeuLeuGlyProGlyAlaGlyAlaGlyGlyThrGlyGlyAl 1085
953 GATGGGCGGAAGGCTGGCGGCACGACTGTGGGGCGGCGGACCTGGCG 1002
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1085 aGlyGlyGlyAlaProAlaAlaProProCysPheAlaAlaProProp 1102
1003 CACTG.....CCGTAGCACTCGCTGGGAGCAGCAGCATCGG 1040
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1102 roCysPheTyrLeuAspValAspGlnSerProSerAspSerGluAspSer 1118
1041 CGGCTGCTTTCAGCAGGTGGGAACCTGACAGCGCCAGGCTGATTGCTGG 1090
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1119 .....GluSerLeuAlaGly.....AlaSerLeuAlaG 1128
1091 CTGGGCGGCCAGCACTTGCTCTGACGGCTTGACCTGGCGCGCCCTCAAGC 1140
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1128 yLeuAspPro.....TrpTrpPheAlaAspPheP 1138
1141 CGCAGCGC..... 1148
1138 roTyrProTyrAlaAspArgLeuGly**ProAlaAlaArgTyrGlyLeu 1154
1149 .....GAGGCTCTCTGCTACTGCGCCAC 1174
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1155 ValAspLysLeuGlyGlyTyrLeuAlaGlySerTrpAspTyrLeuPro** 1171
1175 TGGGACCAAAACCTGCAGCTACACCTGC.....TGTGCTACAGCCAGC 1218
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1171 *ArgSerGlyArgAlaAlaTrpHisCysArgHisCysAlaSer..... 1185
1219 GAACGTGTGAAGGCTGCCCGCTGCTTGGGCTGCGCGCCAGCCCAACA 1268
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1186 .....LeuGluLeuLeuProProArg 1193
1269 CACTCAAGTCACTGCGCGCCAGGCTGGCTCTTGGTGTGGGAAAGTGT 1318
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1194 HisLeuSerCysSerHisAspGlyLeuAspGly..... 1204
1319 AGGCTGGTGGCGAGCTGTCGCCACTGCTTCTTACTCCCTCCCTCGGAGC. 1367
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1205 .GlyTrpTrpAlaProProPro.....ProProTrpAlaAla 1216
1368 .....CCTCTTGGCCCAAAAGTGTGCTGCTGC.....TCTCTC 1406
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1216 laGlyProLeuProArgArgArgAlaArgCysGlyCysProArgSerHis 1232
1407 CCTCTCTCTCCACCCACTCACACTCCCTCCCTCCATCTCTGAGCTCCCT 1454
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1233 ProHisArgProArgAlaSerHisArgThrProAlaAlaAlaAlaPro 1248

```

```

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.US-08-486-273A-58

```


1186LeuGluLeuProProArg 1193
 1269 CACTCAAGTCACGCGCCAGCGCTGCTGTGCTGGAAAGTGT 1318
 1194 HisLeuSerCysSerHisAspGlyLeuAspGly..... 1204
 1319 AGCTGGTGCACCGCTGCTCCCGCTGCTTCTACTCCCTCCCTGGAGC. 1367
 1205 .GlyTrpTrpAlaProPro.....ProProTrpAla 1216
 1368CCTCTTGGCCACAAAGGCTGCTGTC.....TCTCTC 1406
 1216 laGlyProLeuProArgArgArgAlaArgCysGlyCysProArgSerHis 1232
 1407 CTTCTCTCCACCCACCTACCTCCCTCCATCTCTGAGTCCCT 1454
 1233 ProHisArgProArgAlaSerHisArgThrProAlaAlaAlaPro 1248

seq_name: /cgn2_6/ptodata/1/laa/6A_COMB.pep:US-08-940-086A-58

documentation_block:
 uence 58, Application US/08940086A
 ent No. 6111091

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.
 APPLICANT: Ellis, Steven B.
 APPLICANT: Liaw, Chen W.
 APPLICANT: Lu, Chin-Chun
 TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
 NUMBER OF SEQUENCES: 63
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Heller Erman White & McAuliffe
 STREET: 4250 Executive Square, 7th Floor
 CITY: La Jolla
 STATE: CA
 COUNTRY: USA
 ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/940.086A
 FILING DATE: 29-SEPT-97
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/231.193
 FILING DATE: 20-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/052.449
 FILING DATE: 20-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Seidman, Stephanie
 REGISTRATION NUMBER: 33,779
 REFERENCE/DOCKET NUMBER: 24735-9383C
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 450-8400
 TELEFAX: (619) 450-8499
 INFORMATION FOR SEQ ID NO: 58:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1336 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-940-086A-58

alignment_scores:

Quality: 191.00 Length: 450
 Ratio: 1.194 Gaps: 24

Percent Similarity: 35.556 Percent Identity: 24.667

alignment_block:

US-09-786-136-4 x US-08-940-086A-58 ..

Align seg 1/1 to: US-08-940-086A-58 from: 1 to: 1336

353 GGGGCGCTGACAGTGTGACGTACAGCCCGCTGAGCCCCCAGGAACCCC 402
 |||||
 892 GlyMetTyrSerCysSerAlaGluAlaProProAlaLysPr 908
 403 TCCTGCTCTCACAGCCATCTAGATGACGCTCTCTGCCCCACTGTTGCT 452
 |||||
 908 oProPro.....ProProGlnProLeuProSer..... 917
 453 ACTCTAAAGAGCTTCAGAGAGCTCTCCAGCAGCGCTGGGACTCGGAC 502
 917 917
 503 CAAATGGCAACTCCAAAGCGTCTTTGGGGACCCCTGCCACTCCACGGA 552
 |||||
 918ProAlaTyrProAlaPr 923
 553 AGGTCTCGCGCCCGCCACT.....CCCGCTCTCCCTGTCGGC 590
 |||||
 923 oGlyProAlaProGlyProAlaProPheValProArgGluArgAlaSerV 940
 591 TCAGCATGGCGCGCTCACAACTCTGTAGCCCTGACCCGGCGCAGGCC 640
 |||||
 940 alAlaArgTrpArgArgProGlyAlaGlyPro....ProGlyGlyAla 955
 641 AACACACTGAAGCAGCAGCATGACTACTGCTGCTG..... 679
 |||||
 956 GlyLeuAlaAspGlyPheHisArgTyrTyrGlyProIleGluProGlnI 972
 680GTGGCGTGCAGGAGGACGACCTCTCTCGCTGGACGCGC 722
 |||||
 972 yLeuGlyLeuGlyGlyGluAla.ArgAla..... 982
 723 CTTCTGCGCGCTGTGCGCATCTCGCAGGAGCGCTCTGCGTGCCTCG 772
 |||||
 983AlaProArgGlyAlaAlaGlyArg..... 990
 773 CCCTTTGCGTGCACCGCGCGCTGACCCCGCTGCGCTCATTTGCC 822
 |||||
 991ProLeuSerProAlaAlaGlnProProGlnLysPro...P 1004
 823 CTGGGCACTTCTGGGCGCTGCTGTTGGCCAGGGTGACAGTG..... 863
 |||||
 1004 roAlaSerTyrPheAlaIleValArgAspLysGluProAlaGluProPro 1020
 864CCACGGCTCTGCGAGCGCTGCGCGCTGCGCTGCGCTCG 1035
 |||||
 1021 AlaGlyAlaPheProGlyPheProSerProProAla.....ProProAl 1035
 893 TCGGGCTCTGAGG..... 905
 |||||
 1035 aAlaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052
 906GGCCAGCGCCAGGTGGACA..... 923
 |||||
 1052 spGluSerProProAlaProAlaArgTrpProArgSerAspProGluSer 1068
 924GTGGCGGGGCTCAGACACCGGAGCGCTCC 952
 |||||
 1069 GlnProLeuLeuGlyProGlyAlaGlyGlyAlaGlyThrGlyGlyAl 1085
 953 GATGGGCGGAAGGCTGGCGGACCGACCTGCGGGCGCGGCTGGGC 1002
 |||||
 1085 aGlyGlyGlyAlaProAlaAlaProProCysPheAlaAlaProProp 1102
 1003 CACTG.....CCGTAGCACTGCTGCTGGAGCACCAGTGGC 1040
 |||||

864CCAGGCTCTTCAGCCCGTCCCCCAGC 892
1021 AlaGlyAlaPheProGlyPheProSerProProAla.....ProProAl 1035
893 TCGGGCTCTGAGG..... 905
1035 aAlaAlaAlaThAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052
906GCCAGGCCAGGTGGACA..... 923
1052 spGluSerProProAlaProAlaArgTrpProArgSerAspProGluSer 1068
924GTGGGGGGCTTCAGACCCGAGGCTCC 952
1069 GlnProLeuLeuGlyProGlyAlaGlyAlaGlyAlaGlyThrGlyGlyAl 1085
953 GATGGGGGGAAGGCTGGCGCCAGCCAGCTGGCGGGCGGACCTGGGC 1002
1085 aGlyGlyGlyAlaProAlaAlaProProProCysPheAlaAlaProPro 1102
003 CACTG.....CCGTAGCACTCGCTGGCTGGAGCACCGATGCG 1040
1102 roCysPheTyrLeuAspValaspGlnSerProSerAspSerGluAspSer 1118
1041 CCGCTGCTTCGAGCAGTGGGAAGTGCACAGCGCCCAAGGCTGATTGCTGG 1090
1119GluSerLeuAlaGly.....AlaSerLeuAlaG 1128
1091 CTGGGGCCCGCAGCACTTCGCTGAGCGGCTTGACCTGGCGCCCTCAAGGC 1140
1128 yLeuAspPro.....TrpTrpPheAlaAspPhe 1138
1141 CGCAGGCC..... 1148
1138 roTyrProTyrAlaAspArgLeuGly**ProAlaAlaArgTyrGlyLeu 1154
1149GAGGGCTCTTCCTGCTACTGCGCCAC 1174
1155 ValAspLysLeuGlyGlyTrpLeuAlaGlySerTrpAspTyrLeuPro** 1171
1175 TGGGACCAAACTGCAGCTPACACTGC.....TGTGTACAGCCAGC 1218
1171 *ArgSerGlyArgAlaAlaTrpHisCysArgHisCysAlaSer..... 1185
1219 GAACGTGTGAAGCTGCCCTCTGCTGGCTGGCGCCGCCCAACCA 1268
1186LeuGluLeuLeuProProArg 1193
1269 CACTCAAGTCACTGCCGCCCGCAGGCTGCTCTGTGTGGGAAAGTGT 1318
1194 HisLeuSerCysSerHisAspGlyLeuAspGly..... 1204
19 AGGTGGTGGCAGCTGCTCCCTGCTCTTCTTACCTCCCTCCCTGGAGC. 1367
1205 .GlyTrpTrpAlaProProPro.....ProProTrpAla 1216
1368CCTCTTGGCCCCACAAAAGTGCCTGCTGC.....TCTCTC 1406
1216 laGlyProLeuProArgArgAlaArgCysGlyCysProArgSerHis 1232
1407 CCCTCTCTCCACCCCACTACAGCTCCCTCCATCTCTGAGGCTCCCT 1454
1233 ProHisArgProArgAlaSerHisArgTrpProAlaAlaAlaPro 1248

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:us-08-316-650-12
seq_documentation_block:
; Sequence 12, Application US/08316650
; Patent No. 5942496
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Roessler, Blake J.

APPLICANT: Goldstein, Steven A.
APPLICANT: Lin, Wushan
TITLE OF INVENTION: METHODS AND COMPOSITIONS
TITLE OF INVENTION: FOR STIMULATING BONE CELLS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 30-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1442 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-316-650-12

alignment_scores:
Quality: 189.50 Length: 566
Ratio: 0.746 Gaps: 34
Percent Similarity: 44.876 Percent Identity: 26.325

alignment_block:
US-09-786-136-4/rev x US-08-316-650-12 ..

Align seg 1/1 to: US-08-316-650-12 from: 1 to: 1442

1455 CAGGAGCTCAGAGGATGAGGGGAGTGAGTGGGTGGGAGGAGGG 1406
341 GluGlyAlaGlnGlySerArgGlyGluProGlyAsnProGlySerProG 357
1405 AGAGAGCAGCAGCAGCAGCTTTTGTGGGGCAAGAGGGCTCCAGGGAGG 1356
357 yPro.AlaglyAlaSerGlyAsnProGlyThrAspGlyIleProGlyAla 373
1355 GAGTAAAGACAGTGGGGACAGGCTGGCAGCAGCTACACTTTCCACAGC 1306
374 LysGlySerAlaGlyAlaProGlyIleAlaGlyAlaProGlyPheProG 390
1305 ACCAAGAGCCAGCCCTGGCGGCAGTGTGAGTGTGTTGGGTGGGGC 1256
390 yProArgGlyProPro.....Glyp 397
1255 GCCAG.....CCCAAGCAGCAGGGG..... 1236
397 roGlnGlyAlaThrGlyProLeuGlyProLysGlyGlnAlaGlyCluPro 413
1235GCAGCCTTCACACAGTTCGCTGGGC.....TGTAGCAC 1204

414 GlyIleAlaGlyPheLysGlyAspGlnGlyProLysGlyGluThrGlyPr 430
1203 AGCAGGTCTAGCTACAGGTTTGGTCCAGTGGC.....G 1169
430 oAlaGlyProGlnGlyAlaProGlyProAlaGlyGluGlnGlyLysArg 447
1168 CAGTAGCAGGAGAGCCCTCCGGCTGGCGCTTGGAGGCGCCAGGTCAA 1119
447 yAlaArgGlyGluProGlyGlyAlaGlyProIleGlyProGlyGlu 463
1118GGCGCTCAGGCAAGTCTGG 1099
464 ArgGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaG 480
098 GCCC..... 1095
480 yProLysGlyAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProL 497
1094 ..GCAGCCAGCAATCAGCTTGGCGCTGTCTCAGTTCCTCCACTGTCGAAG 1047
497 yGlyAlaAsnGlyAspPro.GlyArgProGlyGluProGlyLeuProG 513
1046 CAGCGCGCATCGGTGCCAGCAGGAGGTGTACGGCAGTGGCCAG 997
513 yAlaArgGlyLeuThrGlyArgProGlyAspAla.....G 525
996 GTCCGGCCCCAGAGTCGGTCCGCCAGCCCTTCCGCCCATCGGAGGC 947
525 yProGlnGlyLysValGly.....ProSerGlyAla 535
946 CTCGGTGTCTGAGCCCCCAGTGTCTCCAGCTGGCTGGCCCTCAGAGC 897
536 ProGly.GluAspGlyArg.....ProGlyProGlyP 547
896 CGAGCTGGGGAGCGGCTGAGGAGGCGGTGGCACTCTCACCCTGGCCA 847
547 roGlnGlyAlaArgGlyGlnProGlyValMetGlyPhe..... 559
846 ACACAGGCGCCAGAGTGGCCAGGGCAATGAGCCGAGGTGGGCTGAG 797
560 ...ProGlyProLysGlyAlaAsnGlyGluProGlyLysAlaGly..... 573
796 GCTGGCGCGGTGCCAGGCAAGGCGGAGCGCGC.....AGA 759
574GluLysGlyLeuAla.GlyAlaProGlyLeuArg 584
758 GCGCTCTCTGAGATGGCGCACAGCGGCGCAGAGGCGCGCTCCAGGCGG 709
585 GlyLeuProGlyLysAspGlyGluThrGlyAlaAlaGlyProProGly.. 600
708 AAGGAGCTGTGTCTCTGAGCGCGCACCAAGGCGAGGTAGTATGTC 659
601ProSerGlyProAlaGlyGluArgGlyGlnGlyAla 613
658 GCTGCTTCACTGTGTGGCTGGCGCTGCGCTCAGGCTACAGATTGT 609
613 la.....ProGlyPro..SerGlyPheGlnGlyLeuP 623
608 GAGCGCGCGCTGAGCGCGCACAGGAGGAGCGGAGGTGGCGCGGA 559
623 roGlyProProGlyProGlyGluGlyGlyLysGlnGlyAspGlnGly 639
558 GGACCTTCGTGGAGTGGCAGGTTCCTCC.....AAGCAGCGCTT 518
640 ilePro.....GlyGluAlaGlyAlaProGlyLeuValGlyProArgG 654
517 GGAGTGGCAATTTGGTCGAGTCCAGCGCGCTCTGGAGAGCTCTCT 468
654 yGluArgGlyPheProGlyGluArgGlySerProGlyAlaGlnGlyLeuG 671
467 GAAGTCTTTAGAGTAGACACAGTGGCGAGGAGCTGATCTAGATGG 418
:|||||

671 lnglyPro.....ArgGlyLeuPro..... 677
417 CTGTGAGAGCAGGAGGGTCTCTGGG.....GCCTCAGC 383
678GlyThrProGlyThrAspGlyProLysGlyAlaAl 689
382 GGCTCTGAGCTGCACACCTGAGGCGCCAGCAGGACCTCCTAGTAG 333
689 aglyProAsp...Gly.ProProGlyAlaGln...GlyPro..... 700
332 TAGCATCCACAGCTACAGGCAAGTGAAGCAGGAGGCGGCGAGAGTGC 283
701ProGlyLeuGlnGlyMetProGlyGluArg.GlyAlaAlaGlyI 715
282 TTACTTGTCTGAAGGGCTCGAAGCCA..... 257
715 leAlaGlyProLysGlyAspArgGlyAspValGlyGluLysGlyProGlu 731
256GCACC 252
732 GlyAlaProGlyLysAspGlyGlyArgGlyLeuThrGlyProIleGlyPr 748
251 TCCGGGCGATGGCTGTGGAACCAAGAGCAGTCTGTGGAGCCCGCCACCA. 203
748 oProGlyProAlaGlyAlaAsnGlyGluLysGlyGluAlaGlyProProG 765
202CCCACCTCTACAGGCGCTCATAGAAAAAATTCAGTG 164
765 lyProSerGlySerThrGlyAlaArgGlyAlaProGlyGluProGlyGlu 781
163 TTCTCCCGAGCTTCTCTCTATCGACTTTAACAGAGTAGATTGCCTGC 114
782 ThrGlyProGlyProAla.....GlyPheAl 791
113 TAGCGCCAGAACCAAGCTCAGACTCCCACTTTGACAAAGATGATCAGG 64
791 aglyProProGlyAlaAspGlyGlnProGly...AlaLysGlyAspGln 807
63 CTGAATCTTCAACCCATAGCAGCAAGAGCCCGCC 29
807 yGluAlaGlyGlnLysGlyAspAlaGlyAlaPro 818
seq_name: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:PCT-US95-02251-12

seq_documentation_block:

; Sequence 12, Application PC/TUS9502251

GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE

; NUMBER OF INVENTION: CELLS

; TITLE OF INVENTION: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: United States of America

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version

; SOFTWARE: #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/02251

; FILING DATE: CONCURRENTLY HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/316,650

; FILING DATE: 30-SEP-1994

; CLASSIFICATION:

; APPLICATION NUMBER: US 08/199,780

251 TCGGGCATGGCTGTGGAAACCAAGGAGCAGTCTGTGGAGCCCCCACCACCA. 203
||||| ||| : : : : : ||| : : : : : ||| |||
748 OProGlyProAlaGlyAlaAsnGlyGluLysGlyGluAlaGlyProProG 765
202CCCACCCCTCTACAGAGGGCTCATAGAAAAAATTGCAGTG 164
||| ||| ||| : : : : :
765 lyProSerGlySerThrGlyAlaArgGlyAlaProGlyGluProGlyGlu 781
163 TTCTCCCGCAGCTTCTCCCTCTATCGACTTTAACAGAGTAGATTGCCCTTGC 114
:
782 ThrGlyProProGlyProAla.....GlyPheAl 791
113 TAGGCCCGCAGACCAAGCTCAGACTCCCGCACTTTTGACAGAGTGATCAGG 64
| ||| :
791 aglyProProGlyAlaAspGlyGlnProGly...AlaLysGlyAspGlnG 807
63 CTGAATCTTCAACCCCATAGCAGCAGCAAGAGCCGCC 29
:
807 lyGluAlaGlyGlnLysGlyAspAlaGlyAlaPro 818

